VOLUME 38 MARCH, 1952 NUMBER 3

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
OF THE
UNITED STATES OF AMERICA

EDITORIAL BOARD

LINUS PAULING, *Chairman* EDWIN B. WILSON, *Managing Editor*
ALEXANDER WETMORE, *Home Secretary* ROGER ADAMS, *Foreign Secretary*
WILLIAM W. RUBEY, *Chairman of the National Research Council*

J. W. BEAMS
R. E. CLELAND
W. W. RUBEY
HARLOW SHAPLEY
F. R. TERMAN
G. W. BEADLE
E. A. DOISY
G. C. EVANS
C. H. GRAHAM
J. T. PATTERSON
GREGORY BREIT
L. C. DUNN
C. S. MARVEL
J. VON NEUMANN
W. H. TALIAFERRO

Publication Office: Mack Printing Company, Easton, Pa.
Editorial Office and Home Office of the Academy:
2101 Constitution Avenue, Washington 25, D. C.

Entered as Second-Class matter, June 21, 1929, at the Post-office at Easton, Pennsylvania, under the Act of August 24, 1912. Accepted for mailing at a special rate of postage provided for in section 1103, act of October 3, 1917. Authorized June 19, 1929.

Published Monthly Price per Annum, $7.50

OFFICERS OF THE ACADEMY

PRESIDENT ..... D. W. BROWN
VICE PRESIDENT ..... EDWIN B. WILSON
FOREIGN SECRETARY ..... ROGER ADAMS
HOME SECRETARY ..... ALEXANDER WETMORE
TREASURER ..... W. J. ROBBINS

INFORMATION TO SUBSCRIBERS

Checks covering subscriptions to the PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES at the rate of $7.50 per annum should be made payable to the National Academy of Sciences, and sent to the Treasurer, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C. Single numbers $0.75. These prices effective January, 1951.

Microfilm copies of the PROCEEDINGS are available to subscribers to the paper edition. The microfilm is a positive, furnished on metal reels, suitably labeled and is distributed at the end of the volume year. Inquiries concerning purchase should be directed to University Microfilms, 313 N. First St., Ann Arbor, Michigan.

Past publications of the NATIONAL ACADEMY OF SCIENCES from 1863 to 1926 and of the NATIONAL RESEARCH COUNCIL from 1916 to 1926 are listed in the Index to the First Ten Volumes of these PROCEEDINGS. This Index appeared as Part 2 of Volume 13, No. 1, January, 1927, and is for sale separately at the price of $1.50. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

Past volumes of the PROCEEDINGS may be obtained at five dollars per volume unless the sale of the volume would break a complete set of Volumes I to XXX. Single numbers may be obtained for fifty cents except where the sale of such numbers would break up a complete volume. Only forty complete sets of the PROCEEDINGS are available for sale—Volumes I to XXX price $200.00. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

THE NATIONAL ACADEMY OF SCIENCES-NATIONAL RESEARCH COUNCIL also issues technical publications reviewing and correlating recent research. The *Bulletin* and *Reprint and Circular Series* have been replaced by a single publication series, available individually, or by subscription at 25% discount. For a list of available publications, with prices and ordering procedures, address the Publications Office, National Academy of Sciences-National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 38 March 15, 1952 Number 3

*CHEMICAL VALENCE FORCES AND BINDING ENERGY CALCULATIONS**

By Charles P. Mueller and Henry Eyring

Chemistry Department, University of Utah

Communicated January 18, 1952

The principal problem of theoretical chemistry is the elucidation of the causes of molecular stability. The first successful explanation of this stability was the first-order perturbational atomic orbital treatment of Heitler and London.

This was followed by the development of a quite different method by Hund, Mulliken and J. Lennard-Jones, the method of molecular orbitals. It should perhaps be emphasized at this point that this is also only a first approximation.

The principal explanation of chemical valency forces, according to the atomic orbital treatment of Heitler and London, arises from the indistinguishability of similar particles in quantal space. The eigenfunctions describing the molecular system are composed of atomic orbitals which retain their original spherical symmetry.

On the other hand, molecular orbital theory ascribes chemical valency forces to an extreme distortion of the individual one-electron eigenfunctions, the product of which is the molecular eigenfunction.

Depending on our choice of approximate molecular wave functions we get two different interpretations of chemical valence theory.

The atomic orbital method is more in accord with classical valence theory, and each molecule does not pose a special problem, as it does in molecular orbital theory. The method also has the advantage that it appears to be precisely what it is, namely, a mere first approximation. The chief disadvantage consists in its complexity.

Using molecular orbital theory each molecule is a more or less special problem, not only quantitatively, but qualitatively. The general procedure for such calculations requires that we assign various orbitals to the

electrons in such a way that each orbital belongs to some irreducible representation of the point symmetry group of the molecule. The net result is a high degree of orthogonality between the various orbitals and a resultant simplification calculationally. At the same time, such orthogonality places an unnecessary restriction on the orbitals. Theoretically such orthogonality is only required between eigenfunctions representing different eigenvalues of the entire molecular system.

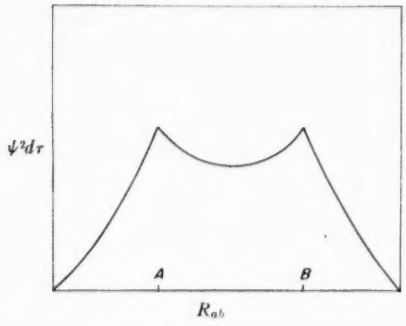

FIGURE 1

Shape of electronic distribution of the one-electron eigenfunctions in a molecular orbital treatment of the hydrogen molecule.

While both of the preceding methods have yielded qualitatively information of great value, they have not been very successful in calculating the things of primary interest to the chemist, namely, binding energies and the observable phase of charge distributions, dipole moments. It seems imperative, therefore, that we pass from first to second approximations, using a self-consistent field method.

As a first approximation to the solution of the Fock-Hartree equation we shall use an eigenfunction capable of assuming molecular or atomic orbital forms as special cases.

We shall apply this method to the hydrogen molecule and then generalize the results to more complicated cases.

The Self-Consistent Field Applied to the Hydrogen Molecule.—Using the method of Fock[1] we find the following set of equations for the hydrogen molecule:

$$\frac{1}{1+S^2}\left[\left\{\int \varphi_i(2)H\varphi_i(2)d\tau_1\right\}\varphi_i(1) + \left\{\int \frac{(\varphi_j(2))^2}{r_{12}}\,d\tau_2\right\}\varphi_i(1) + \left\{\int \frac{\varphi_i(2)\varphi_j(2)}{r_{12}}\,d\tau_2\right\}\varphi_j(1) + \left\{\int \varphi_i(2)H\varphi_j(2)d\tau_1\right\}\varphi_j(1)\right] = \epsilon_i\varphi_i(1); \quad (1)$$

ϵ_i is the energy of the ith electron in the coulombic and exchange field of the other electron.

The correct antisymmetric eigenfunction for the ground state of the hydrogen molecule has the form

$$\Psi = \frac{1}{\sqrt{2+2S^2}}\left\{\varphi_i(1)\varphi_j(2) + \varphi_i(2)\varphi_j(1)\right\}\left\{\alpha(1)\beta(2) - \alpha(2)\beta(1)\right\}. \quad (2)$$

As a first approximation we use the eigenfunction

$$\Psi = \frac{1}{\sqrt{2 + 2S^2}} \{(\varphi_a(1) + \lambda\varphi_b(1))(\lambda\varphi_a(2) + \varphi_b(2)) + (\varphi_a(2) + \lambda\varphi_b(2))(\lambda\varphi_a(1) + \varphi_b(1))\} \quad (3)$$

where φ_a and φ_b are hydrogenic 1s orbitals. This eigenfunction was first used by Coulson and Fischer.[2]

We can then take advantage of a transformation to elliptic coordinates

$$e^{-Zr_a} = e^{\frac{-ZR}{2}(\mu + \nu)}$$

$$e^{-Zr_b} = e^{\frac{-ZR}{2}(\mu - \nu)} \quad (4)$$

where

$$\mu = \frac{r_a + r_b}{R} \qquad \nu = \frac{r_a - r_b}{R}$$

to pass to the second approximation.

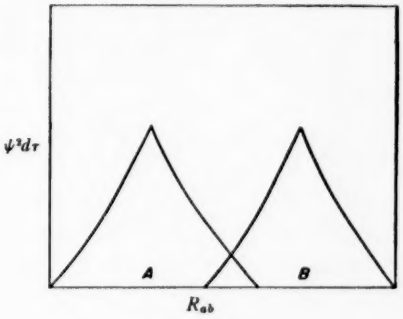

FIGURE 2

Shape of the electronic distribution of the two one-electron eigenfunctions in an atomic orbital treatment of the hydrogen molecule.

In the simple transformation, apart from the sign, the coefficients of μ and ν are equal. By destroying this equality we replace the atomic orbitals by elliptical eigenfunctions

$$e^{-Zr_a} \rightarrow e^{-(\alpha\mu + \beta\nu)} \qquad e^{-Zr_b} \rightarrow e^{-\alpha\mu + \beta\nu} \quad (5)$$

and the eigenfunction of equation (3) becomes

$$\Psi = \frac{1}{N\sqrt{2 + 2S^2}} \{e^{-\alpha(\mu_1 + \mu_2)} [(e^{-\beta\nu_1} + \lambda e^{\beta\nu_1})(\lambda e^{-\beta\nu_2} + e^{\beta\nu_2}) + (e^{-\beta\nu_2} + \lambda e^{\beta\nu_2})(\lambda e^{-\beta\nu_1} + e^{\beta\nu_1})]\} \quad (6)$$

where N is the normalizing factor.

Applying the self-consistent field method we find that $\alpha = 0.835$, $\beta = 0.775$, $\lambda = 0.137$. Such an eigenfunction gives a binding energy of 4.20 e.v. or 89.0 per cent of that observed experimentally. A comparison of the resultant accuracy is given in table 1.

The first approximation to the solution of the Fock-Hartree equation has been called an antisymmetric molecular orbital. For reasons which will appear later we shall refer to this first approximation, the eigenfunction of equation (6) and subsequent generalizations as semi-localized orbitals. Briefly the reason for this is that it is not yet clear that in complicated molecules even the first approximation related to the eigenfunction of

equation (3) will involve the atomic orbitals of all the atoms in the molecule. Secondly, the formalism of the method of semi-localized orbitals is precisely that of atomic orbitals and the final molecular eigenfunction has the symmetry properties of molecular orbitals only for the extreme case of exact coincidence with molecular orbitals.

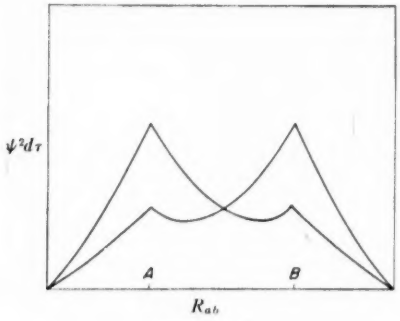

FIGURE 3

Shape of the electronic distribution of the two one-electron eigenfunctions in a semilocalized orbital treatment of the hydrogen molecule.

The charge distributions of the one-electron eigenfunctions of atomic orbital, molecular orbital and semi-localized orbital methods are given in figures 1, 2 and 3.

Extension of the Method to Complex Molecules.—The extension of the preceding method to more complex molecules is made through the use of the formalism previously developed for the atomic orbital method. Each atomic orbital is replaced by a linear combination of elliptic eigenfunctions

$$\varphi_a \rightarrow e^{-\alpha\mu} e^{-\beta\nu}$$
$$\varphi_b = e^{-\gamma\mu} e^{+\delta\nu} \qquad (8)$$

or even more simply by the analogous combination of atomic orbitals.

TABLE 1

HYDROGEN MOLECULE

METHOD	BINDING ENERGY E.V.	BINDING ENERGY KCAL.	TOTAL ENERGY E.V.	TOTAL ENERGY KCAL.
A.O. Variation[a]	3.76	79.66	30.965	96.1
S.C.F.M.O.[b]	3.603	76.27	30.808	96.79
SCF Semi-localized Orbital	4.202	89.02	31.407	98.38

[a] Wang, S., *Phys. Rev.*, **31,** 579 (1928).
[b] Coulson, C. A., *Proc. Camb. Phil. Soc.*, **34,** 204 (1938).

As an example of the simplest sort of eigenfunction, we will report briefly on the application of the method to diatomic hydrogen fluoride. Here we have utilized the eigenfunctions

$$(\varphi_a + \lambda_1\varphi_b); \qquad (\lambda_2\varphi_a + \varphi_b). \qquad (9)$$

Using a complete Hamiltonian, neglecting multiple exchange integrals, and omitting all canonical structures except the one corresponding to

classical valence theory, binding energy and dipole moment calculations were made. Since the molecular orbital and atomic orbital calculations are special cases of semi-localized orbitals, these calculations were made to serve as comparisons. The results are given in table 2 and the calculation will be reported in more detail in a subsequent publication.

TABLE 2

DIATOMIC HYDROGEN FLUORIDE

METHOD	BINDING ENERGY IN A.U.	DIPOLE MOMENT IN D.
Atomic Orbital	−0.9688	0.86
Molecular Orbital	0.1352	0.67
Semi-localized	0.1648	1.56
Observed	0.2440	1.91

In this case the semi-localized orbitals are much closer to the molecular than the atomic orbital treatments. The contrast in dipole moments between all three methods is very noteworthy. The final form for the three treatments is given below.

Atomic Orbital:

$$\varphi_a = \frac{\delta^5}{\sqrt{\pi}} e^{-\alpha r_a} r_a \cos \varphi_a \qquad \delta = 2.60 \varphi_b = \frac{1}{\sqrt{\pi}} e^{-\nu_b}.$$

Molecular Orbital:

$$\varphi_a + 0.571 \varphi_b.$$

Semi-localized Orbitals:

$$(\varphi_a + 0.14 \varphi_b); \ (1.04 \varphi_a + \varphi_b).$$

In this calculation the inner electrons are assumed to be unchanged. This assumption is borne out by the small amount of distortion which the fluorine atomic orbital undergoes. This distortion is assisted by a large exchange force. In the case of the other non-bonding electrons the distortion is opposed by slightly weaker exchange forces. If the valence electron orbital of fluorine is only slightly distorted, it seems likely that the non-valence electron orbitals are practically unchanged.

Calculations with the semi-localized orbitals have now been made for two extreme cases of chemical binding. In both cases considerable improvement in numerical results has been obtained. Several general conclusions seem to be in order. Indiscriminate use of either molecular or atomic orbital methods may lead to nonsensical results, for example, the atomic orbital calculation of the binding energy of hydrogen fluoride. In any domain the semi-localized orbital method is necessarily superior. In the two cases considered the molecular eigenfunction belongs to an irreducible representation of the molecular point symmetry group. The

individual electronic eigenfunctions do not by themselves belong to any such irreducible representation. Hence the molecular orbital formalism is not correct for either molecule considered. It seems likely that this formalism may hold only for the trivial case of a completely ionic bond. The basis for such a conclusion is that for the highly ionic bond of hydrogen fluoride the identity of semi-localized and molecular orbitals is far from being complete.

In any case the exploitation of the method of semi-localized orbitals appears to be a profitable undertaking.

* This paper was prepared in connection with a conference on "Quantum-Mechanical Methods in Valence Theory" held at Shelter Island, September 7–10, 1951, under the auspices of the NATIONAL ACADEMY OF SCIENCES (see "Summary of the Conference" in a subsequent issue of these PROCEEDINGS).

[1] Fock, V., *Z. Physik*, **61**, 126 (1930).

[2] Coulson, C. A., and Fischer, I., *Phil. Mag.*, **40**, 393 (1949).

---

*SUPERPOSITION OF CONFIGURATIONS: THE HELIUM ATOM**

BY G. RUSSELL TAYLOR† AND ROBERT G. PARR

DEPARTMENT OF CHEMISTRY, CARNEGIE INSTITUTE OF TECHNOLOGY, PITTSBURGH, PENNSYLVANIA

Communicated by R. S. Mulliken, January 7, 1952

Introduction.—A very nearly exact approximation to the ground state wave function for the helium atom was obtained by Hylleraas in 1930.[1] His approximation is so accurate that calculation of the ground state wave function and energy for helium no longer remains an interesting problem *per se*, but the extension of his method beyond two-electron problems is quite hopeless.[2] The work reported in the present paper therefore was undertaken as a step in the development of a systematic, practical procedure for obtaining an accurate wave function for any atom or molecule.

The quantitative shortcomings of the methods that have been most commonly applied to many-electron systems heretofore, the Hartree-Fock method for atoms and molecules and the valence bond and molecular orbital methods for molecules, are becoming more apparent with each new application. There has resulted an increased use of wave functions involving configuration interaction.[3, 4] Wave functions of any desired accuracy in principle can be obtained by the superposition of a sufficiently large number of configurations,[5] but if the method is to be of practical value the number of configurations required should be small; i.e., the configurations should be chosen with discretion. Present knowledge re-

garding configuration interaction effects is very meager, however, so that inspiration now serves in place of discretion. Prospects of progress have been termed melancholy,[6] but melancholy or not there appears to be no other road. At the beginning lies the simplest non-trivial example, the helium atom.

In the problem of the helium atom the number of simplifying assumptions is a minimum, the required integrals can be evaluated to any desired accuracy without prohibitive effort, and accurate experimental data are available for evaluation of results. Furthermore the relative simplicity of the system permits the variational determination of the best effective charges to be used in one-electron functions, which is not feasible for most more complex systems. Simplicity, of course, has as its concomitant disadvantage the loss of generality. Nevertheless, conclusions reached for the helium atom can be used as a guide in further exploratory investigations of successively more complex systems, which might have as their goal the formulation of a set (probably a large set, accompanied by massive tables of definite integrals) of empirical rules whereby an accurate configuration interaction wave function for any atom or molecule would be obtainable with a minimum of trial and error effort.

Approximate Ground State Wave Functions and Energies.—We seek a normalized approximate wave function for the ground state of helium which has the form

$$\Psi = \sum_j a_j \Psi_j, \tag{1}$$

where the Ψ_j are normalized antisymmetrized product wave functions (AP's) for various configurations. The best set of coefficients a_j is determined by minimization of the energy $E = \int \Psi^* H \Psi dv$, where H is the Hamiltonian operator for the helium atom neglecting spin-orbit interaction and relativistic effects and assuming infinite mass for the nucleus, namely,

$$H = -\nabla_1^2 - \nabla_2^2 - 4/r_1 - 4/r_2 + 2/r_{12}, \tag{2}$$

where the unit of energy is the atomic unit $e^2/2a_0 = 13.603$ e.v. The best coefficients satisfy the equations

$$\sum_j a_j (H_{ij} - S_{ij}E) = 0, \qquad i = 1, 2, \ldots, \tag{3}$$

and the corresponding best energy E is the lowest root of the secular equation

$$|H_{ij} - S_{ij}E| = 0, \tag{4}$$

where

$$H_{ij} = \int \Psi_i^* H \Psi_j dv \quad \text{and} \quad S_{ij} = \int \Psi_i^* \Psi_j dv. \tag{5}$$

The one-electron functions used in the construction of the AP's are taken as products of the form

$$(i) = (n_i l_i m_i)^{Z n_i}(\eta_i), \tag{6}$$

where $(n_i l_i m_i)^{Z n_i}$ is the hydrogen-like atomic orbital of total quantum number n_i, azimuthal quantum number l_i, magnetic quantum number m_i and effective charge Z_{n_i}, and (η_i) is a spin function (either α or β).

A *configuration* of an atom is defined by the specification of the total and azimuthal quantum numbers for each electron in the system. Belonging to a given configuration there are in general many different AP's corresponding to the many possible assignments of magnetic and spin quantum numbers. We could in equation (1) include all the AP's for each configuration which we wish to superpose and proceed to solve equation (4) by brute force. Since we are neglecting spin-orbit interactions, however, we know in advance that the secular equation may be factored into a product of independent equations, each factor of which involves interactions only between those LS vector-coupled functions belonging to the several configurations which have the same eigenvalues for L^2, S^2, L_z, S_z.[7] Thus in the LS coupling scheme the ^1S ground state wave function for helium, Ψ, is a superposition of ^1S wave functions, Φ_j, belonging to the several configurations:

$$\Psi = \sum_j b_j \Phi_j. \tag{7}$$

The Φ_j for a given configuration are linear combinations of the Ψ_j for that configuration; the procedure for obtaining the former from the latter is well known.[7]

Let us designate the ^1S vector coupled function belonging to the configuration $(1s)^{Z_1}(1s)^{Z_1'}$ by $\Phi(1s^{Z_1}, 1s^{Z_1'})$, that belonging to the configuration $(2p)^{Z_2}(2p)^{Z_2'}$ by $\Phi(2p^{Z_2}, 2p^{Z_2'})$, etc. The approximate wave functions for the ground state of the helium atom which we have investigated may then be written as follows:

$$\Psi_A = \sum_{n_1=1}^{3} \sum_{n_2=1}^{3} A_{n_1 n_2} \Phi(n_1 s^Z, n_2 s^Z), \tag{8}$$

$$\Psi_B = B_1\Phi(1s^{Z_1}, 1s^{Z_1'}) + B_2\Phi(2p^{Z_2}, 2p^{Z_2'}) + B_3\Phi(3d^{Z_3}, 3d^{Z_3'}) + B_4\Phi(4f^{Z_4}, 4f^{Z_4'}), \tag{9}$$

$$\Phi_C = C_1\Phi(1s^{Z_1}, 1s^{Z_1'}) + C_2\Phi(2p^{Z_2}, 2p^{Z_2'}) + C_3\Phi(3d^{Z_3}, 3d^{Z_3'}) + C_4\Phi(4f^{Z_4}, 4f^{Z_4'}). \tag{10}$$

Table 1 gives numerical results of variational calculations using these functions. For a given function, one assumes values for the effective charges, computes the matrix elements H_{ij} and S_{ij} (this requires only well-known procedures[7]), solves equation (4) for the energy E, and repeats with new values for the effective charges until a minimum E is obtained.

Discussion.—A single AP (e.g., function A-1 of Table 1) can be a solution of the wave equation for an atom only if the r_{ij} terms in the Hamiltonian operator are neglected or replaced by a spherically symmetric potential

function. The principal difficulty with such wave functions is that they allow the electrons to move independently of one another; hence, in general, the functions are not small, as they should be, when the electrons are close together. One says that the wave function does not *correlate* the motion of the electrons, and the difference between the correct ground-

TABLE 1

VARIATION FUNCTIONS FOR THE NORMAL HELIUM ATOM

FUNCTION[a]	CONFIGURATIONS SUPERPOSED	PARAMETER VALUES	$-E_{calc.}$ (ATOMIC UNITS[e])	$E_{calc.} - E_{exptl.}$[f, g] (ATOMIC UNITS[e])
A-1 = B-1[b]	$(1s)^2$	$Z = 1.6875$	5.69531	0.1113
A-2	$(1s)^2$, $(1s2s)$	$Z = 1.757$	5.7085	0.0981
A-3	$(1s)^2$, $(1s2s)$, $(1s3s)$	$Z = 1.777$	5.7117	0.0949
A-4	$(1s)^2$, $(1s2s)$, $(2s)^2$	$Z = 1.78$	5.7095	0.0971
A-5	$(1s)^2$, $(1s2s)$, $(1s3s)$, $(2s)^2$, $(2s3s)$, $(3s)^2$	$Z = 1.764$	5.7148	0.0918
B-2	$(1s)^2$, $(2p)^2$	$Z_1 = 1.6896$, $Z_2 = 3Z_1$	5.73870	0.0679
B-3	$(1s)^2$, $(2p)^2$, $(3d)^2$	Z_1 and Z_2 as above, $Z_3 = 6Z_1$	5.742867	0.0637
B-4	$(1s)^2$, $(2p)^2$, $(3d)^2$, $(4f)^2$	Z_1, Z_2, Z_3 as above, $Z_4 = 12Z_1$	5.743770	0.0628
C-1[c]	$(1s1s')$	$Z_1 = 1.19$, $Z_1' = 1.835Z_1$	5.751316	0.0552
C-2	$(1s1s')$, $(2p)^2$	Z_1 and Z_1' as above, $Z_2 = 4.16Z_1$	5.790369	0.0162
C-3	$(1s1s')$, $(2p)^2$, $(3d)^2$	Z_1, Z_1', Z_2 as above, $Z_3 = 9.5Z_1$	5.794033	0.0125
C-4	$(1s1s')$, $(2p)^2$, $(3d)^2$, $(4f)^2$	Z_1, Z_1', Z_2, Z_3 as above, $Z_4 = 16.0Z_1$	5.79486	0.0117
Hartree-Fock[d]	...	...	5.734	0.073

[a] Functions labeled A, B, C are of the types Ψ_A, Ψ_B, Ψ_C of equations (8), (9) and (10), respectively.

[b] Kellner, G. W., *Z. Physik*, **44**, 91 (1927).

[c] Eckart, C., *Phys. Rev.*, **36**, 878 (1930).

[d] Bethe, H., *Handbuch der Physik*, Vol. XXIV/1, p. 324.

[e] One atomic unit $= e^2/2a_0 = 13.603$ e.v. More strictly the values of $E_{calc.}$ should be interpreted as in units of $e^2/2a_{He}$, where $a_{He}/a_0 = 0.999863$, which corrects for the finite mass of the nucleus.

[f] $E_{exptl.} = -5.80656e^2/2a_0$—see Paschen, F., *Sitzber. preuss. Akad. Wiss.*, 1929, p. 662.

[g] These values are calculated assuming those in the previous column are in units of $e^2/2a_0$. Strictly the correction indicated in note *e* should be applied, which increases all $E_{calc.}$-$E_{exptl.}$ values by 0.0008.

state energy of a system and that calculated with the best function of this type (the Hartree-Fock function) is called the *correlation energy*. Thus the correlation energy for the helium atom is 0.073 atomic unit or 0.99 e.-v. (cf. Table 1).

There are essentially two ways (other than explicit introduction of r_{12} in the wave function) in which electron correlation can be brought into an analytical wave function for helium: either the electrons may be separated into concentric spherical shells or the wave function may be given angular dependence (e.g., one electron may be concentrated along the x axis and one along the y axis). In Table 1, all functions in set A and function C-1 fall in the first category, all functions in set B in the second category, and functions C-2, C-3 and C-4 in both categories. Set A separates the electrons into the concentric so-called K, L and M shells, as the same effective charge is used in each of the hydrogen-like orbitals $1s$, $2s$, $3s$. Essentially the same effect is achieved in C-1 through variation of the charges in $1s$ orbitals. It might appear as though such an effect

TABLE 2

EFFECTIVE CHARGES: BEST VALUES VERSUS VALUES DETERMINED BY CRITERION OF MAXIMUM OVERLAP[a]

		Z_1[e]	Z_1'	Z_2	Z_3	Z_4
Function B-4[b]	Best[d]	1.6896	...	$3Z_1$	$6Z_1$	$12Z_1$
	Max. overlap	1.6875	...	$(10/3)Z_1$	$7Z_1$	$12Z_1$
Function C-4[c]	Best	1.19	$1.835Z_1$	$4.16Z_1$	$9.5Z_1$	$16.0Z_1$
	Max. overlap	1.19	$1.835Z_1$	$4.57Z_1$	$9.6Z_1$	$16.6Z_1$

[a] Criterion of maximum overlap: to determine the effective charge for a given high energy configuration wave function, maximize the three-dimensional overlap of the radial part of this function with the wave function for the lowest energy configuration.

[b] See equation (9) and table 1.

[c] See equation (10) and table 1.

[d] Variation of the effective charges was not carried out as precisely with the B functions as with the C functions.

[e] Best and maximum overlap Z_1 values in the case of the B function differ because the former is an adjusted best value for the four-term function B-4 while the latter is the best value for the one-term function B-1. Adjustment does not give appreciable improvement in the case of the C function.

is also operative in the B set of functions, but there it is minor. For although hydrogen-like orbitals for K, L, M and N shells are used, the effective charges are varied, and the best values as given in Table 1 result in almost maximum overlap of the radial portions of the higher energy configuration wave functions with the $1s^2$ wave function (see below). Both effects are operative in the functions C-2, C-3 and C-4. Here $(1s^{Z_1}, 1s^{Z_1'})$ separates the electrons into two shells, and the angular correlation is obtained from $2p^2$, $3d^2$ and $4f^2$ configurations, where again the radial portions of the higher energy functions have almost maximum overlap with the lower energy function.

In Table 2 the variationally determined effective charges for the higher energy configurations in the type B and C functions are compared with

values determined by a maximum overlap criterion. The agreement is striking, particularly in view of the fact that the calculated energies are not very sensitive to small changes in the parameters near their best values.[8]

The angular and radial correlation effects are almost independent of each other. Thus, the improvement resulting from superposition of $2p^2$, $3d^2$ and $4f^2$ configurations is about the same in both B and C sets of functions, even though the radial correlation obtained with B-1 and C-1 are very different (zero in the former case). Successive improvements from introduction of angular factors in the two cases are, in atomic units, 0.0434, 0.0042, 0.0009 for set B and 0.0390, 0.0037 and 0.0008 for set C. On this basis the angular correlation appears to account for about the fraction 0.045/0.073 or 62% of the total correlation energy, radial correlation for about 38%. The best function in Table 1, C-4, contains the fraction 0.0643/0.073 or 88% of the total correlation energy, about 64% of the radial correlation energy, and about 97% of the angular correlation energy. For this function, $C_1:C_2:C_3:C_4::1000:-60:-12:-4$.

Summary.—The wave function for the ground state of the helium atom is approximated by a superposition of vector-coupled ^1S functions for the configurations $(1s^{Z_1}, 1s^{Z_1'})$, $(2p^{Z_2})^2$, $(3d^{Z_3})^2$, $(4d^{Z_4})^2$, where the one-electron functions are hydrogen-like orbitals with the indicated charges. The best calculated energy, 5.79486 $e^2/2a_0$ (experimental value: 5.80656 $e^2/2a_0$), obtained with $Z_1 = 1.19$, $Z_1' = 1.835Z_1$, $Z_2 = 4.16Z_1$, $Z_3 = 9.5Z_1$, $Z_4 = 16.0Z_1$, contains 88% of the total correlation energy (Hartree-Fock energy-experimental energy).

Comparison with similar calculations putting $Z_1 = Z_1'$ suggests that total electronic correlation energy in an atom may be considered as a sum of radial and angular correlation energies. *Radial correlation energy* is tentatively defined as the difference between the Hartree-Fock energy and the energy calculated using the best superposition of all configurations obtainable from a complete set of purely radial one-electron functions (e.g., all *ns* hydrogen-like functions), and *angular correlation energy* is defined as the difference between total and radial correlation energies. The calculations then indicate that the correlation energy in the helium atom is about 62% angular and 38% radial, and that the best wave function obtained contains 97% of the angular and 64% of the radial correlation.

In attempting to regain radial correlation (separate electrons into concentric shells) weakly overlapping shells are found to effect little improvement; hence the use of the complete subset of *ns* functions with only one charge would be a slowly convergent process. Similarly in regaining angular correlation, maximizing the overlap of the *radial* factors of the high-energy angle-dependent terms in the wave function with the purely radial first term is found to yield almost optimum values for the charges.

* This paper was prepared in connection with a conference on "Quantum-Mechanical Methods in Valence Theory" held at Shelter Island, September 7–10, 1951, under the auspices of the NATIONAL ACADEMY OF SCIENCES (see "Summary of the Conference" in a subsequent issue of these PROCEEDINGS). The paper is based on part of a thesis submitted by G. Russell Taylor in partial fulfillment of the requirements for the degree of Doctor of Science, Carnegie Institute of Technology, 1951. The work was assisted in part by the Office of Naval Research, U. S. Navy, through Contract Nonr-493(00) with Carnegie Institute of Technology, in part by a grant from the Research Corporation. G. F. Hadley aided in the computations.

† DuPont Postgraduate Fellow, 1950–1951; Research Corporation Fellow, 1951. Present address: Mellon Institute of Industrial Research, Pittsburgh, Pennsylvania.

[1] Hylleraas, E., *Z. Physik*, **65**, 209 (1930).

[2] James, H. M., and Coolidge, A. S., *Phys. Rev.*, **49**, 688 (1936).

[3] For complex organic molecules, see, for example, Parr, Craig and Ross, *J. Chem. Phys.*, **18**, 1561 (1950) or Coulson, Craig and Jacobs, *Proc. Roy. Soc.*, **206A**, 297 (1951).

[4] For important developments in theory, clear statements of the general method of attack and a calculation on the beryllium atom, see Boys, S. F., *Proc. Roy. Soc.*, **200A**, 542 (1950); **201A**, 125 (1950).

[5] This has long been known and is implicit in statements of many authors, yet correct statements still crop up in the literature. Thus the LCAO molecular orbital method including configuration interaction is identical with the valence bond method including ionic structures, and as the number of AO's is increased both methods converge to the true wave function.

[6] Coulson, Craig and Jacobs, *op. cit.*

[7] Condon, E. U., and Shortley, G. H., *Theory of Atomic Spectra*, Cambridge University Press, Cambridge, 1935; Ufford, C. W., and Shortley, G. H., *Phys. Rev.*, **42**, 167 (1932).

[8] Only after the present work was complete, indeed only after this paragraph was written, was it noticed that S. F. Boys (*op. cit.*, paper II) had earlier also come to the conclusion that maximum overlap of radial functions gives maximum gain of angular correlation energy.

---

*A COMPARATIVE SURVEY OF APPROXIMATE GROUND STATE WAVE FUNCTIONS OF HELIUM ATOM AND HYDROGEN MOLECULE**

BY R. S. MULLIKEN

DEPARTMENT OF PHYSICS, UNIVERSITY OF CHICAGO, CHICAGO 37, ILLINOIS

Communicated January 7, 1952

For both atoms and molecules, there is a great need to find the simplest and most rapidly convergent succession of steps to improve on the familiar approximate wave functions of electron configuration or SCF (self-consistent-field) type. This may be done, in principle, by using wave functions depending explicitly on interelectronic distances r_{ij} (but this is cumbersome in practice), or by forming linear combinations of wave functions of the electron configuration type (configuration interaction). In following

the latter path, Boys has indicated[1] that it is best to use generalized forms of electron configuration type, rather than forms which could, by themselves, be SCF functions. In the present Note, a comparative survey is made in the light of this idea of various approximate forms which have been studied for the wave functions of the ground states of the He atom and the H_2 molecule.[1] The survey also throws some light on the directions in which further studies might be aimed.

Using the ^1S_g helium atom ground state as an example, assumption of an open-shell electron configuration $(s's'')$ permits better energy minimization than the usual closed-shell configuration $(1s)^2$. Leaving aside spin and normalization factors (as will be done everywhere below), the corresponding wave functions are:

$$\text{For } (1s)^2\text{: } 1s^{(1)}1s^{(2)}\text{; for } (s's'')\text{: } s'^{(1)}s''^{(2)} + s'^{(2)}s''^{(1)}, \qquad (1)$$

where $1s$, s', s'' all refer to AO's (atomic orbitals) of s symmetry. The best $(1s)^2$ function is that using Hartree SCF $1s$ AO's. The best $(s's'')$ function would involve two independent but *not* orthogonal forms s' and s'', to be determined by a generalization of the SCF procedure. The superscripts 1 and 2 in Eq. (1) refer to the two electrons. *Here and below*, symbols of the types $(x)^2$ and (xy) are used as *conventional abbreviations* for electron configurations and also especially *for two*-electron wave functions having forms such as are illustrated by Eq. (1).

The simplest approximations to s' and s'' are $1s$ hydrogenic AO's, $1s^\dagger$ and $1s^{\dagger\dagger}$, as used by Eckart and by Hylleraas, with effective Z values Z' and Z'' determined variationally. Somewhat better would be:

$$s' \approx 1s^\dagger + \lambda 1s^{\dagger\dagger}\text{; } s'' \approx 1s^{\dagger\dagger} + \lambda 1s^\dagger; \qquad (2)$$

a little further improvement would be obtained by replacing the variation parameter λ in s'' by an independent parameter. With s' and s'' chosen as in equations (2), the function $(s's'')$ can be written:

$$(s's'') \approx (1s^\dagger 1s^{\dagger\dagger}) + [2\lambda/(1+\lambda^2)][(1s^\dagger)^2 + (1s^{\dagger\dagger})^2]. \qquad (3)$$

This formulation is of interest not particularly *per se*, but in its analogy to certain approximate expressions for the wave function of H_2 (see below).

If we regard $(1s)^2$ as the "true" ground-state electron configuration for helium, then the use of a better function of the type $(s's'')$ (with s' and s'' two s functions) is a first step in taking "configuration interaction" or "correlation" into account. It corresponds to partly taking into account interaction with excited states of the types (ms, ns) and $(ms)^2$. To take these *fully* into account, one might set up a series of mutually orthogonal forms of the type $(s's'')$ and seek to determine these and their coefficients.[2] After that, further improvements, converging toward the exact wave function, would be sought in the addition of ^1S_g forms of the types $(p'p'')$,

($d'd''$), *etc.*—compare the paper of G. R. Taylor and R. G. Parr[1] elsewhere in this issue for part of these.

If we regard ($s's''$) itself as a generalized SCF form, its use may be said to reduce considerably the magnitude of the correlation energy or configuration interaction. The improved correlation in ($s's''$) is accomplished by letting the two electrons move in considerably different AO's, thus increasing their mean distance. This evidently more than outweighs the unfavorable exchange repulsion effect involved in the symmetric form which ($s's''$) possesses (cf. equation (1)).

Using the MO method, a completely analogous treatment can be given for the $^1\sum^+_g$ ground state of the hydrogen molecule. The ordinary SCF closed-shell configuration is $(1\sigma_g)^2$, where $1\sigma_g$ denotes a BMO (best MO). But the open-shell configuration (σ_g' σ_g''), equivalent to partly considering configuration interaction with excited states of types $(\sigma_g\sigma_g)$ and $(\sigma_g)^2$, must give a considerably better wave function.[3] The next step would be to admix $(1\sigma_u)^2$, or better ($\sigma_u'\sigma_u''$), and so on. (It is of interest that the best $1\sigma_u$ has nearly the form of a united-atom hydrogenic $2p\sigma$, although commonly it is approximated by the LCAO form $(1s_a - 1s_b)$. Finally, other $^1\sum^+_g$ MO-product forms of the types ($\pi_u'\pi_u''$), ($\pi_g'\pi_g''$), ($\delta_g'\delta_g''$), and so on, should be added in due proportions.[4]

Using the AO method, the $^1\sum^+_g$ ground state of H_2 has an electron configuration ($\sigma\bar{\sigma}$) of open-shell type; in rough approximation, σ and $\bar{\sigma}$ are $Z = 1$ AO's $1s_a$ and $1s_b$ of the two H atoms (Heitler-London method). It is seen that use of either of the configurations ($\sigma_g'\sigma_g''$) or ($\sigma\bar{\sigma}$) for H_2 may be regarded as a way of improving on $(1\sigma_g)^2$ by the same type of shell-opening as was used in equation (1) for He. And just as in the latter case, the possibility of improved correlation here is realized by letting the two electrons move in two considerably different orbitals, with a net gain in stability over even BMO $(1\sigma_g)^2$ if the two orbitals are chosen in the best way, in spite of the *per se* unfavorable exchange repulsion effect involved in the symmetric form (cf. equation (1) for ($s's''$)) of the wave function.[5] The ($\sigma\bar{\sigma}$) configuration for H_2, however, differs from ($s's''$) for He and ($\sigma_g'\sigma_g''$) for H_2 in that, regarded as MO's, the orbitals σ and $\bar{\sigma}$ do not belong to irreducible representations of the molecular symmetry group; they are nevertheless expressible as simple linear combinations of proper MO's of the two species σ_g and σ_u.

For the best ($\sigma\bar{\sigma}$) functions, PAO's (polarized AO's) may be used for the axially-symmetric orbitals σ and $\bar{\sigma}$, with $\bar{\sigma}$ always the mirror image of σ on reflection in the mid-plane perpendicular to the molecular axis. To approximate σ and $\bar{\sigma}$ analytically, a possible first step is to add hydrogenic $2p\sigma_a$ and $2p\sigma_b$ (with positive ends directed toward the center of the molecule) to $1s_a$ and $1s_b$, respectively, with separate adjustments of effective Z values for $1s$ and $2p\sigma$ so as to minimize the energy. More generally,

$$\sigma = 1s_a^{(Z)} + \alpha 2p\sigma_a^{(Z')} + \beta 3d\sigma_a^{(Z'')} + \ldots. \qquad (4)$$

Using the expansion (4) or some equivalent form, one might proceed to determine BPAO's σ and $\bar{\sigma}$ using the variation method.[6] This should lead to a wave function distinctly better than the well-known Rosen function. Next, additional terms of the same kind as $(\sigma\bar{\sigma})$ but orthogonal to the latter might be added, analogous to what was suggested for He. To improve the wave function further, the next step would be to add functions[4] of the type $(\pi\bar{\pi})$, $^1\sum^+_g$. These were (implicitly) included by Hirschfelder and Linnett in their recent paper on H_2.[7]

An interesting and instructive alternative to equation (4), although not leading so far toward BPAO's, is[8]

$$\sigma \approx 1s_a^{(Z)} + \lambda 1s_b^{(Z)}; \quad \bar{\sigma} \approx 1s_b^{(Z)} + \lambda 1s_a^{(Z)}. \qquad (5)$$

The approximations of equations (5) for σ and $\bar{\sigma}$ are formally analogous to those of equations (2) for s' and s'' of He. From equations (5), using for $(\sigma\bar{\sigma})$ the analog of equation (1) for $(s's'')$, there follows:

$$(\sigma\bar{\sigma}) \approx (1s_a^Z 1s_b^Z) + [2\lambda/(1+\lambda^2)][(1s_a^Z)^2 + (1s_b^Z)^2]. \qquad (6)$$

This expression, which is formally analogous to equation (3) for He, is seen to be the familiar Heitler-London-Wang-plus-ionic approximation for H_2. From what has been said above, it is clear that this cannot be as good as $(\sigma\bar{\sigma})$ with BPAO's for σ and $\bar{\sigma}$. As is well known, equation (6) can also be recast in the form

$$(1\sigma_g)^2_{\mathrm{LCAO}(Z)} + \gamma 1\sigma_u)^2_{\mathrm{LCAO}(Z)}, \qquad (7)$$

which is simply the (Z-adjusted) LCAO approximation to $(1\sigma_g)^2$ improved by taking into account configuration interaction with Z-adjusted LCAO-approximate $(1\sigma_u)^2$.[9]

It will be noted that the PAO's σ and $\bar{\sigma}$ of equations (5) both reduce to the familiar LCAO approximation to the $1\sigma_g$ MO if $\lambda = 1$. In fact, Coulson and Fischer,[8] arriving at these PAO's from the LCAO-MO point of view, call them unsymmetrical MO's!

An interesting and characteristic feature for all the open-shell symmetric forms $(s's'')$, $(\sigma_g'\sigma_g'')$, $(\sigma\bar{\sigma})$ is that when the two orbitals used are chosen in the best way, so as to give minimum energy, they are *not* orthogonal. For example, as Slater has shown,[10] in the case of $(\sigma\bar{\sigma})$ for H_2, if Löwdin's OAO's (orthogonal AO's) are used for σ and $\bar{\sigma}$, the energy is much worse than if $1s_a$ and $1s_b$ are used, or than for $(1\sigma_g)^2$. In fact, the change from natural AO's $1s_a$ and $1s_b$ to BPAO's for σ and $\bar{\sigma}$ is in the direction of increased non-orthogonality (increased overlap). In all of these cases, however, one *could* make use of orthogonal orbitals in various ways; for example, one could write

TABLE 1

SUCCESSIVE APPROXIMATIONS TO GROUND STATES OF He ATOM AND H_2 MOLECULE[a]

APPROXIMATION			He ATOM (^1Sg) AO Basis	H_2 MOLECULE ($^1\Sigma^+g$)	
	No. of Parameters	No. of Orbitals		MO Basis	PAO Basis
Ia	1	1	$(1s^Z)^2$ C.E. = 1.51ev[b]	$(\sigma_g 1s^Z)^2$: LCAO C.E. = 1.26ev[b]	
I	...	1	$(s)^2$: SCF C.E. = 1.00ev[b]	$(\sigma_g)^2$: SCF C.E. = 1.14ev	
IIa	2 (or 1)	2	$(s^{Z'}s^{Z''})$ C.E. = 0.75ev	$(\sigma_g 1s^{Z'}\sigma_g 1s^{Z''})$	$(1s_a{}^Z 1s_b{}^Z)$ C.E. = 0.97ev[b]
II	...	2	$(s's'')$ or[c] $(\alpha s_1)^2 + (\beta s_2)^2$	$(\sigma_g'\sigma_g'')$ or[c] $\alpha(\sigma_{g1})^2 + \beta(\sigma_{g2})^2$	$(\sigma\bar{\sigma})$ C.E. < 0.59ev[d] (The PAO form II is comparable in generality to the MO form II′.)
II′	...	4 (or 2)		$(\sigma_g'\sigma_g'') + \gamma(\sigma_u'\sigma_u'')$	
III[e]	∞	∞	$\Sigma a_{ij}(s_i s_j)$ or[c] $\Sigma c_i(s_i)^2$ Either form III is equivalent to $e^{-\delta(r_1+r_2)}\ \Sigma c_{pq} r_1{}^p r_2{}^q$	$\Sigma a_{ij}(\sigma_{gi}\sigma_{gj})$ or[c] $\Sigma c_i(\sigma_{gi})^2$	$\Sigma a_i(\sigma_i\bar{\sigma}_i)$
III′	∞	∞		$\Sigma a_{ij}(\sigma_{gi}\sigma_{gj}) + \Sigma b_{ij}(\sigma_{ui}\sigma_{uj})$ or[c] $\Sigma c_i(\sigma_{gi})^2 + \Sigma d_i(\sigma_{ui})^2$	
				The PAO form III and the MO form III′ are equivalent to each other and to the form $e^{-\delta(\xi_1+\xi_2)}\ \Sigma c_{pqrs}\xi_1{}^p\eta_1{}^{2q}\xi_2{}^r\eta_2{}^{2s}$, where ξ, η are elliptical coordinates; C.E. = 0.46ev.	
IV	∞	∞	Linear combinations of III with analogous forms of ^1S_g type constructed using p, d, f, ... AO's. Equivalent to Hylleraas' $e^{-\delta(r_1+r_2)}\ \Sigma c_{pqr} r_1{}^p r_2{}^q r_{12}{}^s$ C.E. = 0.00	Lin. combs. of III′ forms with analogous forms of $^1\Sigma^+g$ type constructed using π_u, π_g, δ_g, ... MO's.[f]	Lin. combs. of III form with analogous forms of $^1\Sigma^+g$ type constructed using π, δ, ... PAO's.[f]
				Equivalent to James and Coolidge's $e^{-\delta(\xi_1+\xi_2)}\ \Sigma c_{pqrst}\xi_1{}^p\eta_1{}^{2q}\xi_2{}^r\eta_2{}^{2s}r_{12}{}^t$ C.E. = 0.00	

[a] C.E. means correlation energy, that is, departure from energy of best wave function.

[b] For best choice of Z's.

[c] See footnote 2.

[d] This is based on Gurnee and Magee's 2-parameter approximation to $(\sigma\bar{\sigma})$—cf. footnote 6.

[e] The terms in each summation are supposed to be taken mutually orthogonal.

[f] Cf. footnote 4.

$$\begin{aligned} s' &= \sum_i c_i' s_i; \quad s'' = \sum_i c_i'' s_i; \\ \sigma_g' &= \sum_i c_i' \sigma_{gi}; \quad \sigma_g'' = \sum_i c_i'' \sigma_{gi}; \\ \sigma &= \sum_i c_i' \sigma_{gi} + \sum_j c_j'' \sigma_{uj}; \quad \bar{\sigma} = \sum_i c_i' \sigma_{gi} - \sum_j c_j'' \sigma_{uj}, \end{aligned} \tag{8}$$

where in each case the forms s_i, σ_{gi}, σ_{uj} are members of a convenient complete orthogonal set of orbitals of the indicated symmetry, with the c''s suitably different from the c'''s. *Or*, for example, supposing s' known, s'' may be written as $s' + \alpha s^*$, where s^* is orthogonal to s'; α and the particular form of s^* (and of s') must be determined by variation methods.

The viewpoints presented above can be summarized and somewhat extended by means of the following table.

* This paper was prepared in connection with a conference on "Quantum-Mechanical Methods in Valence Theory" held at Shelter Island, September 7–10, 1951, under the auspices of the NATIONAL ACADEMY OF SCIENCES (see "Summary of the Conference" in a subsequent issue of these PROCEEDINGS).

[1] This survey Note was inspired by pre-conference reading, including especially S. F. Boys, *Proc. Roy. Soc.*, **201A**, 125 (1950), Kotani's and Taylor and Parr's abstracts in the conference program, and references given below. Kotani and especially Lennard-Jones at the conference presented similar viewpoints as to H_2, in connection with reports on extensive and interesting computations. See also T. Inui, *Proc. Phys.-Math. Soc. Japan*, Sec. 3, **20**, 770 (1938).

[2] The form $(s's'')$ may be rewritten in terms of two mutually orthogonal AO's s_1 and s_2 in place of the non-orthogonal s' and s'', as follows. Let

$$s' = a's_1 + b's_2; \quad s'' = a''s_1 + b''s_2, \tag{A}$$

where, if s', s'', s_1, s_2 are now all taken as normalized,

$$a'^2 + b'^2 = 1 \text{ and } a''^2 + b''^2 = 1.$$

Then, for normalized $(s's'')$,

$$(s's'') = 2a'a''N(s_1)^2 + 2b'b''N(s_2)^2 + (a'b'' + a''b')N(s_1s_2), \tag{B}$$

$$N = [2(1 + Q^2)]^{1/2}; \quad Q \equiv \int s's''\,d\tau, \tag{C}$$

N being the factor which normalizes the form $(s's'')$ of (1), and the forms $(s_1)^2$, $(s_2)^2$ and (s_1s_2) being defined in the same manner as in (1). Equation (B) takes various forms depending on how a' or a'' is chosen in (A). Of interest are the choices $a' = 1$ (hence $b' = 0$) and $a'' = 1$ (hence $b'' = 0$); or especially the choice $a'b'' + a''b' = 0$. The latter gives

$$(s's'') = N\{(1 + Q)(s_2)^2 - (1 - Q)(s_1)^2\}, \tag{D}$$

with

$$s_1 = [2(1 - Q)]^{-1/2}(s' - s''); \; s_2 = [2(1 + Q)]^{-1/2}(s' + s'') \tag{E}$$

Because of this equivalence of the forms $(s's'')$ and $\alpha(s_1)^2 + \beta(s_2)^2$, the same final results should be obtained if one minimizes the energy by varying the latter form based on two orthogonal AO's, or the former based on two non-orthogonal AO's.

[3] Hylleraas, if I remember correctly, once tried such a function, with an H_2^+ BMO (or nearly so) for σ_g', and a similar but lower-Z MO for σ_g''.

[4] The forms of $(\pi'\pi'')$ and the like are somewhat more complicated than that of $(\sigma'\sigma'')$. To obtain a form of ${}^1\Sigma^+_g$ type, one must take

$$(\pi'\pi'') = [f(1)g(2) + f(2)g(1)] \cos(\varphi_1 - \varphi_2),$$

where $\pi' \equiv fe^{i\Phi}$; $\pi'' \equiv ge^{i\varphi}$.

[5] Usually the stability in the AO method of the symmetric form $(1s_a\ 1s_b)$, as compared with $1s_a(1)\ 1s_b(2)$ alone, is said to be due to the occurrence of a *negative* exchange integral instead of the usual positive integral found for atoms. A detailed analysis reveals, however, that the apparent negative exchange energy in $(1s_a\ 1s_b)$ is the sum of a positive exchange energy as for atoms, plus other numerically larger negative terms of the same nature as the *resonance* terms which make H_2^+ stable as compared with H plus H^+ (cf. R. S. Mulliken, *Chem. Rev.*, **9**, 354 (1931)).

[6] An interesting short-cut toward BPAO's is that proposed by Gurnee and Magee (*J. Chem. Phys.*, **18**, 142 (1950)), who took σ as $1s^{(Z)}$ centered about a point somewhat displaced from nucleus a toward nucleus b.

[7] Hirschfelder, J. O., and Linnett, J. W., *J. Chem. Phys.*, **18**, 130 (1950).

[8] Coulson, C. A., and Fischer, I., *Phil. Mag.* [7], **40**, 386 (1949); also Slater, J. C., *Phys. Rev.*, **82**, 538 (1951).

[9] The amount of configuration interaction, i.e., the value of γ, is of course considerably greater here than if BMO's had been used for $1\sigma_g$ and $1\sigma_u$; the same sort of thing is always expected when wave functions are not based on BAO's or BMO's.

[10] Slater, J. C., *J. Chem. Phys.*, **19**, 220 (1951).

THE RADIATION SENSITIVITY OF CATALASE AS A FUNCTION OF TEMPERATURE

By R. B. SETLOW

BIOPHYSICS DIVISION, YALE UNIVERSITY,* NEW HAVEN, CONNECTICUT

Communicated by G. E. Hutchinson; read before the Academy, November 5, 1951

Introduction.—A complete solution to the problem of the direct action of radiation on large molecules necessitates the knowledge of three things: (1) the amount and form of energy lost in passing through the material being investigated, (2) the manner in which this energy is dissipated in the material and (3) the response of the system to the energy put into it. If one uses a type of radiation whose physical effects are well known, then the sensitivity of large molecules to this radiation can give information about the structure of the molecule. If the radiation employed is a beam of fast deuterons, we can be reasonably certain of the first item.[1] The manner in which the energy released in a molecule is dissipated is not well known, but an extension of the present work should shed additional light on this problem. We will be primarily concerned here with the response of an aggregate of catalase molecules to the energy released by the passage of deuterons through them. Charged particle irradiation of large molecules with specific functions usually gives the result that one primary ionization inactivates the entire molecule.[2, 3] However, previous work on the irradiation of catalase by deuterons[4] indicated that an ionization within

the catalase molecule did not destroy all its catalytic activity. Some viruses show results qualitatively similar to those obtained for catalase, and the usual reason given for the failure of an ionization to inactivate them is that the released energy only destroys a part of the structure, and the rest is able to function by itself.[1] In the case of T-1 bacteriophage it has been shown that the radiation sensitivity increases as the bombardment temperature increases.[5] The present work reports somewhat similar results for the irradiation of catalase by deuterons. The results will be analyzed in terms of structure of the catalase molecule.

Experimental Procedure.—In order to eliminate any indirect effects due to a suspending medium, all the irradiations were carried out upon dry catalase. The catalase used was derived from beef red cells.† Samples for bombardment were prepared by pipetting 0.05 ml., containing amounts up to 100 μg. solids, onto small glass coverslips. The coverslips were dried in a vacuum desiccator and then placed around the periphery of a circular copper disk which was put in an insulated bombardment chamber of the cyclotron in which the pressure was $<10^{-4}$ mm. Hg. The disk could be rotated from outside the chamber so as to expose different samples to the deuteron beam. The current carried by the deuteron beam was measured by a galvanometer. The area of the beam was measured from the blackening of photographic film placed in the beam, and the time of exposure was controlled by an electrically operated shutter. Further details may be found elsewhere.[3]

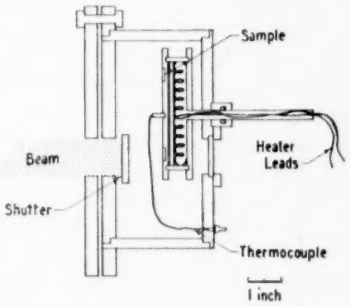

FIGURE 1

The bombardment chamber with provision for raising the temperature during irradiation.

For irradiations above room temperature, the bombardment disk was heated electrically, as shown in figure 1. The temperature during the bombardment was measured by a calibrated chromel-alumel thermocouple. The temperature was kept within $\pm 2°$C. of the average temperature. For bombardments below room temperature a device similar to that shown in figure 1 was used, except that the electrical heater was absent. The inside of the bombardment disk could then be filled with liquids at different temperatures, such as liquid air or a water-ice mixture.

After bombardment the samples were removed from the cyclotron, the catalase dissolved off the coverslips in water and assayed for enzymatic activity by comparing the amount of hydrogen peroxide decomposed in a

known time with a previously constructed calibration curve. The activity of control samples, which underwent the same treatment except for irradiation, was also determined. The hydrogen peroxide was determined either by the amount of I_2 it formed from KI[4] or from its absorption at 247mμ as found in a Beckman spectrophotometer.[6]

If the samples are deposited on quartz disks, it is possible to determine how much material is left on the coverslip after immersion in water by measuring the absorption of the disk in the ultra-violet region of the spectrum. Data on irradiated and control samples indicated that at least 90 per cent of the material was dissolved off.

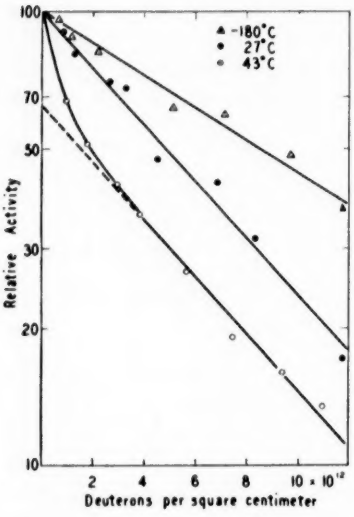

FIGURE 2

A semilog plot of the relative activity of catalase vs. the number of incident deuterons/cm.^{2} for several irradiation temperatures.

The energy of the deuterons used was 3.8 m. e. v. It is possible to calculate from published data[7] that these deuterons, in passing through protein, will lose on the average 235 e. v./100A. It is known that fast charged particles can lose their energy in three ways: by ionization, excitation and nuclear recoils. For 3.8-m. e. v. deuterons only the first two are of significance, and because of the low quantum yield for excitation,[8] we shall neglect the second and consider that the important energy loss item is that due to ionization. It is known from analysis of energy loss by charged particles in gases that the average energy lost per ionization is about 110 e. v. We shall assume the same value holds for solids.[1] This primary ionization can of course give rise to energetic electrons which can produce ionizations of their own.

Results.—Some of the experimental results are shown in figure 2. The relative activity of catalase is plotted on a logarithmic scale against the number of deuterons per cm.^{2}. It is seen that the room temperature curve is a straight line. This indicates that the relative activity may be given by an expression of the form $\frac{A}{A_0} = e^{-SD}$, where A_0 is the activity of the controls, A the activity of the samples irradiated by a dose of D deuterons per cm.^{2} and S is a proportionality constant which we shall call the inac-

tivation cross-section. This expression of course simply indicates that the loss of activity under irradiation is a random affair, just as is radioactive decay. If we want to be literal about it, we can give the proportionality constant S physical significance. S has the units of an area, and thus it can represent the cross-section of a volume which is inactivated by the passage of a deuteron through this volume. For example, from the data taken at room temperature, one can compute that the cross-section corresponds to 1420 A^2. This is seen by finding the dose for which $\frac{A}{A_0} = e^{-1} = 0.368$. The cross-section is then simply the reciprocal of this dose. Since catalase has a molecular weight of 225,000,[9] and thus a cross-sectional area of about 6000 A^2, the inactivation cross-section is obviously too small. But a deuteron can pass through a catalase molecule without losing any energy at all, since the distribution of primary ionizations along a deuteron path is a random one. The chance of such a miss may be found by using bombarding particles of different velocities.[2] Previous work with different energy deuterons and with 2-m. e. v. electrons,[10] which because of their very sparse ionization give an inactivation volume instead of cross-section, indicates that a cross-section of 1420 A^2 corresponds to a volume whose molecular weight (assuming a density of 1.33 g./cm.^{3}) is about 110,000. Thus if we assume catalase has a molecular weight of 225,000, then a primary ionization at room temperature knocks out one-half of its activity. It is seen in figure 2 that irradiation at liquid-air temperature yields a significantly smaller cross-section than at room temperature.

Data taken at 43°C. give a more complex curve than at room temperature. This curve, however, can be resolved into at least two components, one of which has the same cross-section as the room temperature curve. The rapid decline in activity at low doses has been associated with a partial heat inactivation of catalase.[4] It should be pointed out, however, that dry catalase is not appreciably inactivated for several days at a temperature of 43°C. Since the reasons for this initial rapid decline in some of the curves obtained are not well understood at present, we shall neglect these data and concentrate on the part with the small cross-section. Other typical data are shown in figure 3. While cross-sections from these data may be obtained with high precision, it is only fair to recognize that the absolute accuracy is nowhere nearly as good. The major source of this error lies in the possible lack of uniformity in the cross-section of the deuteron beam and in the calibration of the beam galvanometer.

Figure 4 is a compilation of data obtained from runs similar to those already presented. Each point represents one run except for 300° and 90°K. which represent averages of five runs each. The general trend is for the cross-section to increase as the temperature goes up. This increase, however, is not by any means uniform. There definitely seem to

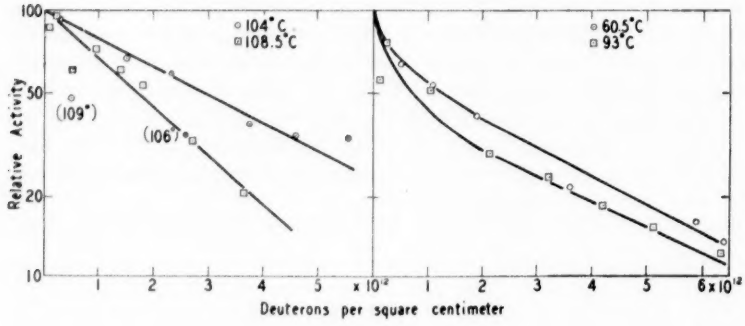

FIGURE 3

A semilog plot of the relative activity of catalase vs. the number of incident deuterons/cm.^{2} for several irradiation temperatures. Note the similar slopes for 60.5° and 93°C. and the big difference between 104° and 108.5°C. The points of the 104°C. run which were at a higher temperature are indicated.

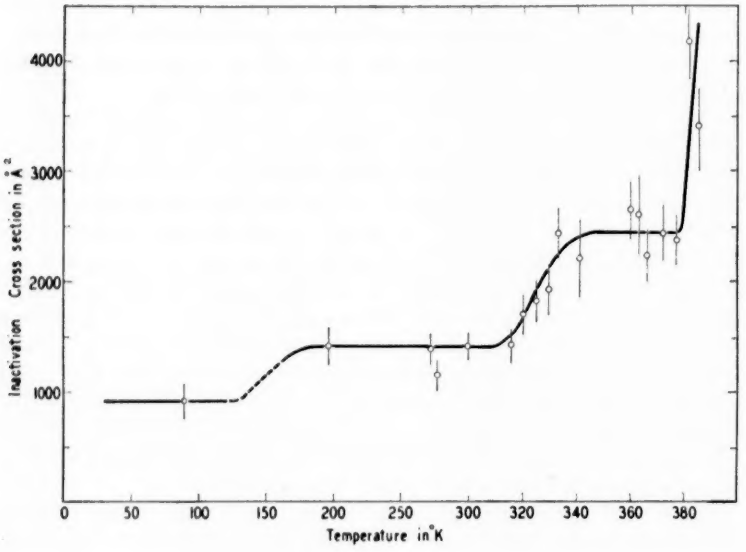

FIGURE 4

The inactivation cross-section for 3.8-m. e. v. deuterons vs. the temperature during irradiation. The temperature scale is not uniform.

be temperature intervals over which the cross-section remains constant. Somewhat similar results have been found for invertase.[11] At the highest temperatures indicated the rate of heat inactivation is appreciable. At 385°K. the specific rate constant for inactivation is about 2×10^{-3} sec.$^{-1}$.[4]

Discussion.—In the analysis that follows we shall assume that the molecular weight of catalase is 225,000. As mentioned previously, we can take other data to indicate that a cross-section of 1420 A^2 corresponds to the inactivation of a molecular weight of 110,000, or just about one-half the molecule, by a primary ionization.

The only thing that we are varying in this experiment is the manner of degradation of the energy released within the catalase molecule by an ionization. Since the average energy released for all the temperatures involved is the same, the relative cross-sections at different temperatures are measures of the sensitivity of the molecule at these different temperatures. Because a molecule is a three-dimensional structure, it is more significant to calculate the volumes associated with these cross-sections. Since the volume depends upon the 3/2 power of the area, it is a simple matter to find the volume inactivated by a primary ionization at various temperatures, from the experimental observation that 1420 A^2 corresponds to a molecular weight of 110,000 inactivated per primary ionization. The results have been put in terms of molecular weights since these are more recognizable than volumes. We thus obtain the values presented in table 1.

TABLE 1

THE MOLECULAR WEIGHT INACTIVATED PER PRIMARY IONIZATION FOR DIFFERENT TEMPERATURES

TEMPERATURE, °K.	CROSS-SECTION, A^2	MOLECULAR WEIGHT PER PRIMARY IONIZATION	MOLECULAR FRACTION PER PRIMARY IONIZATION
300	1420	110,000	½
90	925	58,000	¼
350	2450	250,000	1
385	4200	500,000	2

Qualitatively there is a simple explanation for these results. At liquid-air temperatures all the bonds in the molecules are in their ground vibrational states. Any large amount of localized energy introduced into the molecule will be degraded to thermal energy within a relatively small volume. As the temperature is increased the energy will still fall off rapidly with distance from the site of introduction, but because some of the higher vibrational levels are now excited, it is easier for a new molecular configuration to arise. The data indicate that the catalase molecule can function in parts. If this were not so, the cross-section versus temperature curve would show a steady increase with temperature. Since catalase has

four heme groups, the correlation between the molecular fraction inactivated and parts of the molecule is excellent. It is of interest that there is no evidence for a cross-section plateau corresponding to three-quarters of the molecule. This may mean that a quarter catalase molecule is inactive, as has been suggested,[12] or that inactivation of three-quarters of the molecule alters it so drastically that its entire structure is changed, and it no longer is a closely knit entity. At temperatures above 380°K. the cross-section rises above that corresponding to the whole molecule. At these temperatures the heat inactivation rate is large, and any slight additional energy could cause inactivation of the entire structure. If this energy comes from a deuteron passing outside the molecule, owing, for example, to a secondary electron or an energetic excitation, then obviously one would measure a volume greater than the size of the molecule.

Summary.—The sensitivity of catalase to irradiation by 3.8-m. e. v. deuterons depends on the temperature at which the bombardment is performed. From 90° to 385°K. three different regions of constant sensitivity have been found. They correspond to inactivation of one-quarter, one-half and the entire enzymatic activity of the molecule. There is no evidence for the inactivation of three-quarters of the molecule.

Acknowledgments.—The author wishes to thank Mr. J. W. Preiss for his help in running the cyclotron and Professor E. C. Pollard for many helpful discussions.

* Assisted in part by the U. S. Atomic Energy Commission.

† Armour Powdered Catalase 30 from Armour and Company, Chicago, Ill.

[1] Pollard, E., and Forro, F., *Arch. Biochem. Biophys.*, **32**, 256 (1951).

[2] Lea, D. E., *Actions of Radiations on Living Cells*, Macmillan, New York, 1947.

[3] Pollard, E., *Am. Scientist*, **39**, 99 (1951).

[4] Setlow, R. B., *Arch. Biochem. Biophys.* **34**, 396 (1951).

[5] Pollard, E., and Adams, W. R., *Ibid.*, in press.

[6] Chance, B., and Herbert, D., *Biochem. J.*, **46**, 402 (1950).

[7] Livingston, S., and Bethe, H. A., *Rev. Mod. Phys.*, **9**, 263 (1937).

[8] McLaren, A. D., *Advances in Enzymol.*, **9**, 75 (1949).

[9] Sumner, J. B., Dounce, A. L., and Frampton, V. L., *J. Biol. Chem.*, **136**, 343 (1940).

[10] Slater, M., Ph.D. Thesis, Yale University, 1951.

[11] Pollard, E., Powell, W., and Reaume, S., these PROCEEDINGS, **38**, 173 (1952), immediately following.

[12] Chance, B., *Nature*, **161**, 914 (1948).

*THE PHYSICAL INACTIVATION OF INVERTASE**

BY ERNEST POLLARD, W. F. POWELL AND S. H. REAUME†

BIOPHYSICS DIVISION, SLOANE PHYSICS LABORATORY, YALE UNIVERSITY, NEW HAVEN, CONNECTICUT

Communicated by P. R. Burkholder; read before the Academy, November 5, 1951

Although invertase was one of the first enzymes discovered, many facts concerning its physical properties are still unknown. Conflicting figures are given for its molecular weight[1, 2] and data quoted on the thermal inactivation[3] do not take into account the partial reversibility of inactivation of invertase discovered by Herriott.[4] In this work we have observed the effect of heat on invertase in the dry and wet conditions, of fast electron bombardment, of deuteron bombardment and of combined heat and deuteron bombardment. In addition we have studied the effect of deuteron bombardment on solubility of the enzyme and on the serological activity as measured by combination with specific antibody.

The result has been to enable a measurement of the molecular weight of the unit of invertase which is enzymatically active, of the unit which is serologically active, and to show the relationship between thermal and ionizing radiation action on invertase regarded as a protein.

Thermal Inactivation.—The general stability of invertase to various forms of inactivation was measured roughly in the early days of enzymology.[5] Very little study has been made of dry inactivation. Herriott[4] showed that at pH 4.6 invertase is reversibly inactivated when heated in solution, though the reversibility applied apparently only to part of the enzymatically active material. Chase, Reppert and Ruch[6] showed clearly that inactivation reached an equilibrium at 50°C. and pH 7.

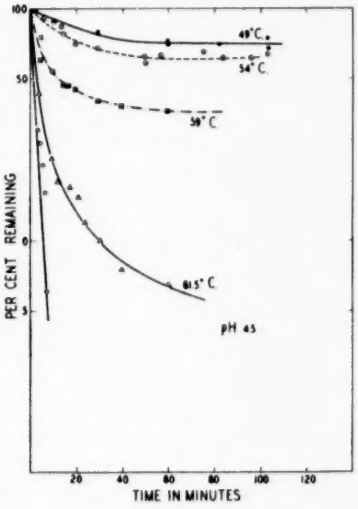

FIGURE 1

The thermal inactivation of wet invertase at pH 4.5. There is clearly a reversible process occurring at low temperatures.

We have studied the effect of temperature at pH 4.5, pH 7 and in the dry state over a range of temperatures. The enzymatic activity was deter-

mined by the method of Sumner and Howell.[7] The enzyme (melibiase-free) was supplied by Nutritional Biochemical, Inc.

Typical results of the study are shown in figure 1. This shows the wet inactivation at pH 4.5. In these curves the per cent activity remaining is plotted as a function of time at various temperatures. It can be seen that at low temperatures the inactivation proceeds to a definite fraction and thereafter remains steady. This is the feature discovered by Chase, Reppert and Ruch.[6] At pH 7 and 41°C. the inactivation remained at 50% after seven hours of exposure. Such samples regain still more activity after cooling to lower temperatures but the process is not completely reversible. This is in agreement with Herriott's findings.

If we treat the process as one of reversible inactivation so that we have a reaction constant k_1 for inactivation of native enzyme and k_2 for the recovery from the inactive form the measurements allow values of k_1 and k_2 to be deduced from the curve. These are shown in table 1.

TABLE 1

pH	TEMPERATURE, °C.	k_1, FRACTION/SEC.	k_2, FRACTION/SEC.	F_{k1}, CAL./MOLE
4.5	48.5	1.7×10^{-4}	4.14×10^{-4}	24,500
	54.0	3.8×10^{-4}	6.8×10^{-4}	24,500
	59.0	1.4×10^{-3}	8.0×10^{-4}	23,950
	61.5	3.26×10^{-3}	2.1×10^{-4}	23,550
	64.0	5.75×10^{-3}	...	23,350
7.2	41.5	4.10×10^{-4}	4.26×10^{-4}	23,550
	44	9.50×10^{-4}	3.90×10^{-4}	23,150
	46.5	1.56×10^{-3}	3.90×10^{-4}	23,000
	49.0	2.78×10^{-3}	...	22,800
	54.0	1.19×10^{-2}	...	22,150
	59.0	3.20×10^{-2}	...	21,900
Dry	90	5.8×10^{-6}	...	30,250
	100	2.04×10^{-5}	...	30,100
	107	5.55×10^{-5}	...	29,900
	117	9.7×10^{-5}	...	30,100
	126	3.30×10^{-4}	...	30,200
	144	1.39×10^{-3}	...	30,250

It can be seen that for both pH 4.5 and pH 7 k_1, the rate of the forward reaction increases markedly with temperature. On the other hand, k_2 first rises and then falls very sharply. For this reason studies of the reactivation of invertase above 65°C. fail to show any appreciable reactivation.

Analyzing the variation of k_1 with temperature in terms of the theory of absolute reaction rates[8,9] we can derive the free energy $\Delta F^{\ddagger}$ of activation. The resulting values are given in table 1 for the forward reaction. It can be seen that $\Delta F^{\ddagger}$ diminishes with temperature. If we put

$$\Delta F^{\ddagger} = \Delta H^{\ddagger} - T\,\Delta S^{\ddagger}$$

we deduce $\Delta H^{\ddagger} = 53{,}000$ cal./mole at pH 7 and $\Delta S^{\ddagger} = 95$ cal./mole per degree while at pH 4.5 the figures are $\Delta H^{\ddagger} = 48{,}500$ and $\Delta S^{\ddagger} = 75$ cal./mole.

These are much lower than those quoted by Glasstone, Laidler and Eyring.[3] We suggest that the discrepancy is due to the existence of the reverse recation at low temperatures.

In the dry state the reaction constants are as shown in table 1. Invertase is very much more stable when dry. We have not found any conclusive evidence for reactivation in the dry state. The temperature variation fits a constant value of $\Delta F^{\ddagger}$ at 30,000 cal./mole. The entropy of activation is accordingly zero or very nearly so. This fits a general behavior already noticed for some bacterial viruses and for two or three enzymes for which it is found that in the dry state $\Delta S^{\ddagger}$ is either zero or slightly negative.[10, 11]

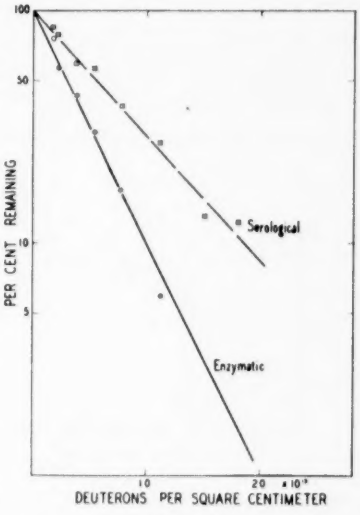

FIGURE 2

The inactivation of invertase by deuteron bombardment. The per cent remaining, both for enzymatic and serological activity, is a logarithmic function of the number of deuterons per square centimeter.

Inactivation of Dry Invertase by Ionizing Radiation.—The fact that ionizing radiation produces a number of separated but highly active agents—the primary ionizations—renders it possible to use this distributed energy to form some estimates of the size and shape of molecules which are sensitive to the action of radiation. This was used by Svedberg and Brohult[12] who studied the splitting of hemocyanin by alpha particles in solution. Further studies on ribonuclease and myosin by Lea, Smith, Holmes and Markham,[13] indicated that very large fractions of the proteins were radiation sensitive. Deuteron and electron bombardment is more precise in that the location of the primary ionizations is definite. This was tried for pepsin and trypsin by Pollard, Buzzell, Jefferys and Forro[14] and reasonably accurate values for the molecular weights of these enzymes found. Setlow[15] found a similar result to hold for urease. Catalase[16] and hemoglobin[17] are apparently not wholly radiation sensitive under normal conditions, a fact which is linked to their multiple structure.

Deuteron and electron bombardment as described by Pollard, Buzzell,

Jefferys and Forro[14] was applied to invertase. It consisted in irradiating dry samples with known numbers of deuterons or electrons. We are indebted to Mr. Marvin Slater for help with the electron irradiation.

The results of deuteron bombardment on enzymatic action are shown in figure 2 (lower line). The relation $\ln n/n_0 = -SD$ (where n/n_0 is the fraction remaining, S the apparent cross-section per deuteron and D the number of deuterons incident per square centimeter) is obeyed. S should be related to the molecular area if the deuteron beam is 100% efficient in inactivation.

To test this, the energy of the deuteron beam was varied and S measured. For very slow deuterons the ionization per unit path is much greater than for fast deuterons and so a greater efficiency should result. The findings are plotted in figure 3 where the observed S is seen to increase with ionization density to a value of 5.1×10^{-13} cm.^{2}. Some correction has to be applied for the fact that at low velocities deuterons more readily produce delta rays. The proportion of these can be estimated from theory[18] and have been tabulated for deuterons by Lea;[19] the estimated correction is shown by the crosses. The area of the sensitive part of the molecule is 4.0×10^{-13} cm.^{2}. The dotted curve is derived from simple theory on the basis that the thickness of the molecule is 50 A and the fit is reasonably good.

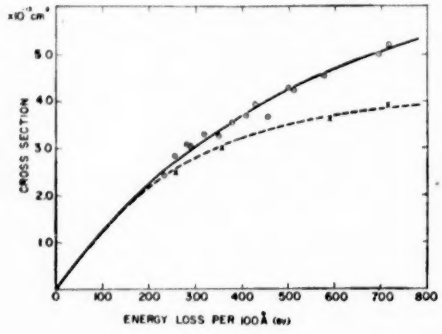

FIGURE 3

The apparent target cross-section for deuteron bombardment measured for various energy deuterons with a variety of values of the energy loss per 100 A. The crosses are corrected for delta ray effects and the dashed line is a theoretical efficiency curve assuming the invertase molecule to be 50 A thick.

The above measurement gives information on the effective *area* of the molecule. If ionization which is random in *volume* is used the inactivation curve takes the form $\ln n/n_0 = -VI$, where I is the number of primary ionizations per unit volume and V is the target volume.

Such volume-random ionization is produced by fast electrons. Accordingly dry enzyme preparations were bombarded by fast electrons from a linear accelerator. The appropriate logarithmic inactivation was found and the mean value for the inactivation volume was 1.52×10^{-19}cm.^{3}.

We thus have three pieces of information about the inactivation of invertase.

(*a*) The target area for perfect efficiency = 4.0×10^{-13} cm.^{2}
(*b*) The target volume is 1.5×10^{-19} cm.
(*c*) The effective thickness is 50 A.

If it is assumed that the invertase molecule is a cylinder of radius r and length l, and that the molecules are lying flat when bombarded, the first two lead to a radius of 24.2 A and a length of 83 A. The third piece of information checks reasonably well with a diameter of 48.4 A.

The use of ionizing radiation in this way leads to the idea that invertase is a somewhat elongated molecule.

The molecular weight, assuming a density of 1.35 is 123,000.

The only available data on the molecular weight of invertase are contained in the diffusion measurements by Euler, Josephson and Myrback[1] which when analyzed by the Einstein-Sutherland relation (a correction suggested by Dr. R. B. Setlow) lead to a value of 120,000 for the dry enzyme.

Serological Assay.—In order to investigate the effect of deuterons on the ability of the molecule to combine with specific antibody, injections of invertase were made into a rabbit, and specific antiserum prepared. We were assisted in this work by Mr. Wilbur J. McNulty. The loss of ability to precipitate antibody is shown in the upper line of figure 2 and it is clear that the cross-section is less than for enzymatic inactivation. The target area found is 1.25×10^{-13}cm.^{2}. The estimated molecular weight of the fraction responsible for serological behavior is 27,000.

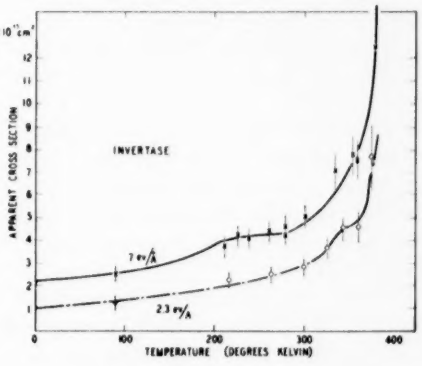

FIGURE 4

The apparent cross-section per deuteron for different temperatures of the specimens bombarded. The enzyme sensitivity increases to a steady value at the cross-section of the molecule and thereafter rises sharply as excitations and photons become able to produce inactivation.

One very important factor in protein inactivation is solubility. To check on this the amount of material which goes into solution after irradiation was measured by weighing the amount recovered after solution and evaporation. Within the accuracy of the experiment all the invertase was soluble after bombardment, unless very heavy doses were given.

Combined Effect of Temperature and Deuteron Bombardment.—The re-

markable behavior of ionizing radiation in removing enzymatic activity from a volume as large as the measured volume of an invertase molecule requires some interpretation. In order to find some basis for this, the effect of keeping invertase preparations at a wide variety of temperatures while irradiating was studied. An effect of combined temperature and ionizing radiation on a bacterial virus in the dry state was found by Adams and Pollard[20] and an effect of heating before and after irradiation has been reported by Setlow for catalase.[16] The results of runs taken from liquid air temperature to 100°C. are shown in figure 4.

Two bombarding energies were used, representing the extremes possible in ionization density. The series at 7 e. v. per Å show three significant facts. The first is that at low temperatures the apparent cross-section is less than that at room temperature. The second is the constant cross-section at 4.2×10^{-13}cm.^{3} from 220° to 290°K., which represents the cross-section for the whole molecule and the third is the considerable upward sweep at high temperatures.

Two runs with the electron accelerator at 300° and 100°K. showed that at 100°K. the inactivation volume is about 0.67 times that at 300°K.

Discussion.—In the dry state, for thermal inactivation, invertase has an activation energy of 30,000 cal./mole with no entropy of activation. This corresponds to the passage over a potential barrier of 1.3 electron volts. No reversibility is observed. This barrier can actually correspond to the breaking of one bond, which seems rather hard to believe in a cross-bonded protein, or to the breaking of several. Three hydrogen bonds are a possibility

Wet inactivation differs in that the molecule is hydrated and therefore subject to the effect of electrostatic forces due to groups which can be ionized. The dissociation of these groups at low pH accounts for the variation of the heat of activation.

The action of ionizing radiation alone appears to be explainable in terms of the requirement that a primary ionization fall within the molecular volume. This releases 110 e. v. and undoubtedly causes the rupture of many bonds. It nevertheless appears as though this effect, large as it may be, does not have a 100% chance of inactivation because at low temperatures there is a clear drop in efficiency. It is possible that the cause of this lies in the differential internal expansion of the protein. Unlike most crystalline solids, proteins are structurally dependent on three different classes of bonds, those in the polypeptide chain, those linking the side chains in a helix and those linking residues between helices. The effect of lowering the temperature may be to cause stronger bonding between residues holding the helices in place. Raising the temperature would cause a weakening.

On this view the inactivation at low temperatures requires more than

the normal energy release at a primary ionization. At higher temperatures where excitation, as opposed to ionization, is sufficiently energetic to cause inactivation the apparent cross-section rapidly rises.

The serological cross-section to deuteron bombardment is definitely lower than the enzymatic cross-section. Two explanations of this can be given. The first is that the serological properties are less specific and therefore that damage can be sustained without removing the ability to combine with antibody. This may, indeed be true but we are inclined to propose a second explanation in terms of the presence of a second non-enzymatically active protein in large excess. The invertase we used did not have either the high activity or the thermolability of preparations described by Sumner and O'Kane[21] and so must contain some protective protein. The serological cross-section would indicate that this protein has a molecular weight of 27,000. Judging from the low effect of deuteron bombardment on solubility this protein must be highly soluble.

Summary.—Thermal inactivations of invertase at pH 4.5 and pH 7 shows that a partly reversible loss of activity occurs with the reverse reaction constant passing through a maximum at 59° for pH 4.5 and 44° for pH 7. For the forward reaction the value of $\Delta H^{\ddagger}$ is 53,000 cal./mole at pH 7 and 48,500 at pH 4.5. Dry invertase shows a different thermal behavior with no observed reversibility and with a value of $\Delta H^{\ddagger}$ of 30,000 cal./mole. There is no entropy of activation in the dry state.

Deuteron and electron bombardments of dry invertase fit with the idea that one primary ionization inactivates the whole molecule. The calculated size of the molecule, assumed cylindrical, is 48 A in diameter and 83 A long. The molecular weight is 123,000.

Combined thermal and deuteron bombardment gives a varying cross-section which shows a constancy over 100°K. at the molecular size but is smaller at low temperatures and much higher at high temperatures. The variable sensitivity is tentatively explained in terms of differential expansion inside the molecule.

Serological studies indicate that the invertase used also contains a protective protein of high solubility and molecular weight 27,000.

Acknowledgments.—We are indebted to Mr. Wilbur J. McNulty for preparing the rabbit antiserum and for carrying out the serological assay. Many bombardments were carried out by Mr. J. W. Preiss, to whom we express our thanks. We owe a great deal to Dr. R. B. Setlow for discussions on both technique and theory.

* Assisted in part by the U. S. Atomic Energy Commission.

† Present address, Biology Department, Stanford University.

[1] Euler, H., Josephson, K., and Myrback, K., *Hoppe-Seyler's Z.*, **130**, 87 (1923).

[2] Moelwyn-Hughes, E. A., *Ergeb. Enzymforsch.*, **2**, 1 (1933).

[3] Glasstone, S., Laidler, K. J., Eyring, H., *The Theory of Rate Processes*, McGraw-Hill, 1941, p. 443.

[4] Herriott, R. M., Ph.D. dissertation, Columbia University, 1932.
[5] Salkowski, E., *Z. Physiol. Chem.*, **31,** 305 (1900–1901); **61,** 124 (1909).
[6] Chase, A. M., Reppert, E. H., and Ruch, R. M., *J. Cell. Comp. Physiol.*, **23,** 27 (1944).
[7] Sumner, J. B., and Howell, S. F., *J. Biol. Chem.*, **108,** 51 (1935).
[8] Eyring, H., *J. Chem. Phys.*, **3,** 107 (1935).
[9] Kunitz, M., *J. Gen. Physiol.*, **32,** 241 (1948).
[10] Pollard, E., and Reaume, M., *Arch. Biochem. Biophys.*, **32,** 278 (1951).
[11] Pollard, E., *Am. Scientist*, **39,** 99 (1951).
[12] Svedberg, The, and Brohult, Sven, *Nature*, **143,** 938 (1939).
[13] Lea, D. E., Smith, K. M., Holmes, B., and Markham, R., *Parasitology*, **36,** 110 (1944).
[14] Pollard, E., Buzzell, A., Jeffreys, A., and Forro, F., Jr., *Arch. Biochem. Biophys.*, **33,** 9 (1951).
[15] Setlow, R. B., *Ibid.*, in course of publication.
[16] Setlow, R. B., *Ibid.*, **34,** 396 (1951).
[17] Appleyard, R. K., *Ibid.*, **35,** 121 (1951).
[18] Bothe, H., *Handbuch der Physik*, **24,** 273 (1933).
[19] Lea, D. E., *Actions of Radiations on Living Cells*, Cambridge, 1947, p. 28.
[20] Adams, W. R., and Pollard, E., *Arch. Biochem. Biophys.*, in course of publication.
[21] Sumner, J. B., and O'Kane, D. J., *Enzymologia*, **12,** 251 (1948).

IONIZING PARTICLE EVIDENCE FOR THE MOLECULAR WEIGHT OF THE PNEUMOCOCCUS TRANSFORMING PRINCIPLE

BY DONALD FLUKE, RUTH DREW AND ERNEST POLLARD

BIOPHYSICS DIVISION, YALE UNIVERSITY,* NEW HAVEN, CONNECTICUT, AND THE MEDICAL AND BIOLOGY DEPARTMENTS, BROOKHAVEN NATIONAL LABORATORY, UPTON, LONG ISLAND

Communicated by P. R. Burkholder; read before the Academy, November 5, 1951

Introduction.—The type specific transformation of pneumococcus R cells to S cells by an extract of S cells was first demonstrated by Griffith[1] for intraperitoneal injection into mice, and later was demonstrated *in vitro* by Dawson and Sia.[2] Avery, MacLeod and McCarty[3] showed that the active transforming principle is in the desoxyribosenucleic acid (DNA) fraction of the S cells. The DNA which they extracted was similar to DNA prepared from other sources and its molecular weight was therefore estimated to be about 500,000. Direct observation of the active DNA in the ultracentrifuge does not provide a determination of the molecular weight, since the boundary diffusion is insufficient to give the necessary information about particle shape.

Our purpose is to report evidence consistent with a larger particle size for the transforming principle. Dry preparations of the active DNA have

been bombarded with 2-m. e. v. electrons and 3.8-m. e. v. deuterons with indication of a molecular weight about 6,000,000.

The first attempt to determine a molecular weight for a biological molecule by ionizing radiation was by Svedberg and Brohult[4] and later Lea, Smith, Holmes and Markham[5] found molecular weights for ribonuclease and myosin. The technique has since been applied to numerous materials by various authors. Pollard and Forro[6] and Pollard, Buzzell, Jeffreys and Forro[7] give a complete discussion of the technique as it has been applied in the work reported here.

The process by which ionizing radiations indicate particle size or structure may be summarized briefly. These radiations in passing through material lose energy to the material in many local events, whose size and spatial distribution may be determined. If the material is dry, the action is direct and can affect the structure of the material only at the locales where the energy is released. Where the material embodies some biological principle, easily destroyed by even minor changes in its structure, the relation of inactivation of the principle to distribution of energy release gives an indication of the size and shape of the particles of the principle.

Deuterons of the energy used in this work distribute ionization events densely along straight paths through the material, inactivating in accordance with the relation $N/N_0 = e^{-SD}$, where N/N_0 is the fraction of activity remaining after subjection to a dose D deuterons/cm.^{2}. The deuterons thereby probe for the frontal area of sensitive regions of cross-section S cm.^{2}/deuteron.

Since the electrons used ionize much more sparsely and do not follow straight paths through material, the incident doses are translated into ionizations/cm.^{3}. The ionizations are distributed randomly in volume, inactivating as $N/N_0 = e^{-VI}$, where I is the number of ionizations per unit volume and V is the volume in which at least one ionization must occur in order for inactivation to result. A molecular weight can be calculated for V, knowing the density of the material. Comparison with the deuteron cross-section gives an indication of the shape of the particles.

Experimental.—The data for this work were obtained in eleven runs with the 4-m. e. v. cyclotron at Yale, two runs with the 2-m. e. v. Van de Graaff machine in the chemistry department at Brookhaven National Laboratory, and one run with the linear accelerator at Yale, also for 2-m. e. v. electrons. The bombardment chambers used with the cyclotron and with the Van de Graaff machine are described in reference 7. The chamber used with the linear accelerator was fully insulated for beam collection, was arranged to collect secondary electrons and opened directly to the accelerator vacuum.

A preparation of active DNA diluted with buffered saline was pipetted in amounts from 5 to 105 micrograms of DNA onto $^1/_2$-inch glass cover slips, ar-

ranged in lots of eight or ten on bombardment disks. The DNA specimens were dried in a refrigerated desiccator pumped on through a liquid air trap.

A smooth strain, pneumococcus Type III (A66) was used as a source of DNA. The active principle was prepared according to the improved method of McCarty and Avery.[8] Strain R36A, a rough variant derived originally from a virulent "s" culture of pneumococcus Type II was used as the test organism for transforming activities remaining for the various DNA samples after exposure to ionizing radiation.

The special environmental conditions necessary for transformation of the susceptible R cells have been discussed in detail by Avery, MacLeod and McCarty[3] and by Taylor.[9] *In vitro* transformation of the unencapsulated R strain to the encapsulated S strain requires not only the addition of the DNA to the nutrient materials but the further addition of an agglutinating agent for the R cells, an accessory protein constituent of serum and sodium pyrophosphate. In the present study, anti-R36A rabbit serum was used as the agglutinating agent and the accessory protein constituent was supplied in the form of a purified fraction bovine albumin.[9] It was found convenient to use a solution of albumin pyrophosphate which was prepared just prior to use by mixing one part of a $M/15$ solution of sodium pyrophosphate with ten parts of a 4% albumin solution. Charcoal-adsorbed heart infusion broth prepared as described by MacLeod and Mirick[10] constituted the assay medium.

After exposure to the beam the cover slips with the DNA specimens were dropped into tubes containing 0.5-ml. buffered saline. This was done as soon as possible, so that the total time dry was usually not over 4 to 6 hours. Tubes containing 8.6 ml. of broth, 0.1 ml. of a 1:10 dilution of antiserum, 0.4 ml. of albumin-pyrophosphate solution and 0.25 ml. of a 10^{-4} dilution of a 6- to 7-hour blood broth culture of the test organism were set up. The specimen cover slips in their tubes were crushed into fine pieces using the tip of a 1-ml. serological pipette and the saline solution of DNA was thoroughly remixed by repeatedly drawing it up and down in the pipette. The solutions were transferred to tubes of the assay medium and the resuspension tubes rinsed with 0.2 ml. of broth. After careful and uniform mixing of the DNA solutions with the assay medium, each mixture was distributed in 2-ml. amounts among five small tubes. The tests were incubated at 37°C. for 18 hours and examined for clouding of the broth which indicates the growth of transformed cells. In each series of five tubes the number in which transformation occurred was noted. To confirm the results a loopful of culture from each tube was streaked onto a section of a blood agar plate. The plate cultures were examined after incubation at 37°C. for 16 hours for the appearance of glistening, mucoid colonies typical of pneumococcus Type III.

The assay procedure, giving as the result some number of positive trans-

formations among five observations, is a sensitive indicator of the amount of transforming principle present only in the neighborhood of some critical amount, corresponding to two or three transformations out of five. The assay does not permit simply putting some DNA on a cover slip and after bombardment removing it to assay the per cent activity remaining. Instead, the same amount of DNA was put on several cover slips (usually four or five), which were then bombarded with different doses. The assay result was then interpolated to find a dose corresponding to survival of the critical amount, which when divided by the amount originally put on the slips gave the relative activity remaining.

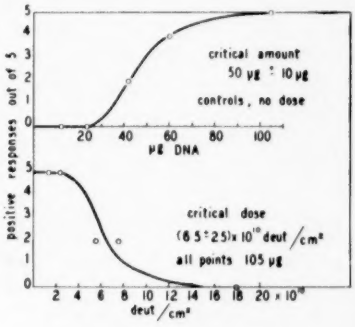

FIGURE 1

Transformations out of five as a function of the amount of DNA and of the dose, illustrating the method of determining per cent activity remaining.

The critical amount was determined for each bombardment by a control series subject to the same experimental conditions throughout, except for actual exposure to the beam. Cover slips with graduated amounts of DNA were fastened to the bombardment discs along with the specimens to be irradiated. Interpolation of their assay results determined the critical amount applicable in that experiment. As an example, figure 1 shows the result of a typical run. The control series amounts were 12, 24, 42, 60 and 105 micrograms of DNA. Taking three out of five positive as the arbitrary end-point, the critical amount indicated is 50 micrograms. The irradiated points all received 105 micrograms of DNA and were exposed to doses of 1.3, 2.4, 5.7, 7.7 and 18 (all $\times 10^{19}$) deuterons/cm.^{2}. Amounts and doses were usually chosen in approximately geometric series, since the critical amount varied considerably from run to run and there was risk of losing the result by attempting to pull the data closer to a

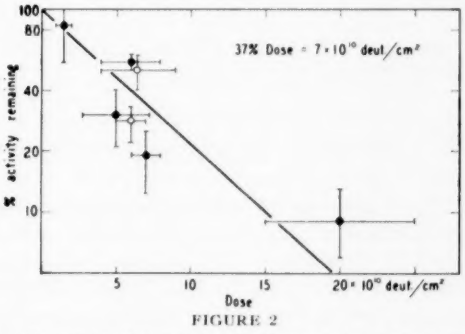

FIGURE 2

Per cent activity remaining (log scale) as a function of deuteron dose.

predicted end-point. At three out of five, the same arbitrary end-point as for the control series, the interpolated dose in figure 1 is 7×10^{10} deuterons/cm.^{2}, corresponding to 50/105 micrograms, or approximately 50% activity remaining.

The data in figure 1 also illustrate the estimation of assay uncertainty, carried over into the dose-effect curves presented below. In the example the critical amount is deemed reliable only between 40 and 60 micrograms (40 to 60% activity remaining) and the dose between 4 and 9×10^{10} deuterons/cm.^{2}. On the dose-effect curves these uncertainties are expressed as limits of error

Results and Interpretation.—The principal data of this experiment are collected in the next two figures. Figure 2 shows the per cent activity remaining as a function of dose in deuterons/cm.^{2}. The error brackets refer only to the uncertainties in interpreting the transformation assay. The two points which are not blackened are the result of runs at approximately doubled ion density (by placing foils in the beam), and it may be observed that there is no change in the cross-section within the dispersion of the data. If the sensitive regions had a dimension less than about 60 to 80 A in the beam direction (or randomly oriented), such a change should be detectable. A straight line has been drawn through the data, corresponding to a cross-section of 1.5×10^{-11} cm.^{2}/deuteron.

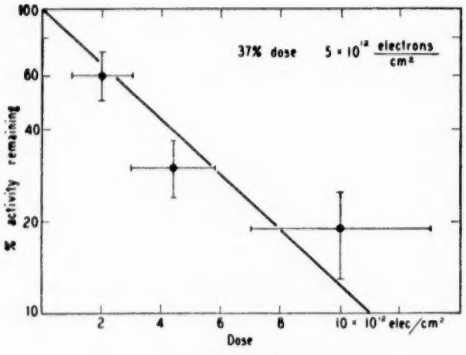

FIGURE 3

Per cent activity remaining (log scale) as a function of electron dose.

Figure 3 shows the result for bombardment with 2-m. e. v. electrons, plotting per cent activity remaining as a function of dose in electrons/cm.^{2}. The 37% dose is 5×10^{12} electrons/cm.^{2}. The point for the largest dose is the result of bombardment with the pulsed beam from the linear accelerator. Between the other two points, obtained with the continuous beam from the Van de Graaff machine, the current differed by a factor of five. The agreement among these three runs is taken as an indication of reciprocity in intensity and time for dose as well as that the other experimental differences involved do not greatly influence the result.

Before analyzing the data one additional series of runs, shown in figure

4, is presented. These data were taken at liquid air temperature and show a very much diminished cross-section for the 3.8-m. e. v. deuterons, about one-third that at room temperature. When the ion density was doubled, however, resulting in the two open circle points, the cross-section increased to larger than that at room temperature.

In these experiments resuspension efficiency was checked by examining the optical densities of the amounts recovered, using a Beckman ultraviolet spectrophotometer. A control and a point irradiated with enough deuterons to have inactivated it to about 10% activity remaining were compared, 60% as much material being recovered from the latter as from the former. There was no change in the ultra-violet spectrum, which was that characteristic of nucleic acid. Our results have been adjusted by correcting the dose-effect curves at 10% activity remaining and extrapolating back to 37%, where a cross-section was chosen between the two curves, assuming that the same proportion of non-resuspended as of resuspended DNA had been inactivated. The correction at 37% amounts to about 10% decrease in cross-section. A similar correction has been made for the electron data.

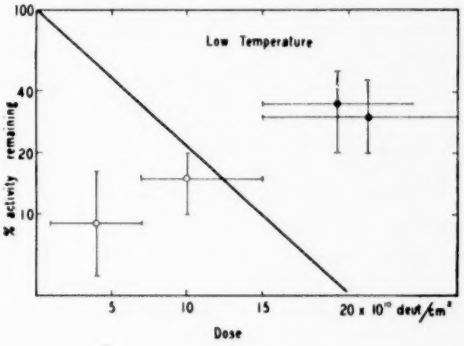

FIGURE 4

Per cent activity remaining (log scale) as a function of deuteron dose at 90°K. The line represents room temperature deuteron inactivation, and the open circle points are for data at doubled ion density at the low temperature.

The primary data presented above, correcting for resuspension, are: (1) deuteron cross-section 1.4×10^{-11} cm.^{2}/deuteron; (2) electron critical dose 5.5×10^{12} electrons/cm.^{2}. The electron critical dose is now multiplied by the energy loss per centimeter for 2-m. e. v. electrons in DNA, taken to have a composition approximately $N_{15}C_{39}C_{24}H_{50}P_4$, on the basis of the DNA base distribution given by Davidson.[11] The density of DNA, 1.65 g./cm.^{3}, was taken from the same reference. From these figures for composition and density the energy loss calculated from the Bethe-Block formula[12] is 3.1 m. e. v./cm. The only appreciable difference from that for protein is due to the increased density. The average energy per primary ionization event is taken as 110 e. v./p. i. Multiplying the electron critical dose per centimeter squared by 3.1×10^6 e. v./cm. and dividing by 110 e. v./p. i. gives 1.5×10^{17} p. i./cm.^{3}. The sensitive volume in

which at least one ionization must occur for inactivation is then 6.5×10^{-18} cm.^{3}/p. i. Multiplying this volume by the density of DNA (which cancels its previous use) and by Avogadro's number gives a molecular weight: 6,500,000.

The fact that the DNA deuteron cross-section is larger than that of a sphere of the above critical volume is in agreement with much evidence[11] that the DNA particles are fibrous, or long thin cylinders of high axial ratio. Dimensions of a cylindrical shape may be obtained by comparing the area and the volume. The most reasonable assumption about the orientation of such cylinders in our work is that they are randomly oriented in direction, since the dried DNA preparations were about 100,000 A thick. For a randomly oriented cylinder of length l and diameter d, the average cross-sectional area (assuming the ends negligible) is $(\pi/4)ld$. Taking $(\pi/4)\ ld = 1.4 \times 10^{-11}$ cm.^{2} and $(\pi/4)ld^2 = 6.5 \times 10^{-18}$ cm.^{3} leads to $d = 45$ A and $l = 3800$ A, giving an axial ratio of about 80.

Discussion.—Our data indicate a sensitive region of molecular weight about 6,000,000 associated with the transforming activity of pneumococcus DNA. At least one ionization must occur in a region of this size for inactivation to result. We may now identify these sensitive regions with the actual particles of DNA, since in order for the particles to have a smaller size it would be necessary to account for their inactivation by radiation-produced events not taking place within the actual structure itself and in the absence of effective means of energy transport.

The molecular weight of nucleic acid has been variously reported,[11] ranging from values corresponding to quite simple polynucleotides up to a molecular weight of 3,000,000 reported by Gulland, Jordan and Threlfall[13] for DNA. In determining the molecular weight of nucleic acid a considerable source of uncertainty has been its possible degradation by the methods required to separate it from protein. As these methods have been refined the trend has been toward higher molecular weights reported for nucleic acid. There has also been uncertainty as to whether aggregation may occur during purification and extraction procedures.

By working with a biological activity of nucleic acid rather than with the physical properties of entire preparations it is possible to by-pass many of these uncertainties. Biologically active DNA is presumably more closely identified with DNA as it actually exists in place in living cells and their fragments. Use of the ionizing particles looks through possible degradation products and even heterogeneous materials to show the size of these active particles. Aggregation may be ruled out also, since activity is not enhanced by the fractionation procedures.

Our results with low-temperature bombardment need further clarification. In particular the low-temperature work should be extended to include electron bombardment. The decrease in deuteron cross-section

at low temperature may mean that the DNA particles are enabled to sustain one ionization within their structures, but no more. Increasing the ion density by a factor of two could then give the observed increase in cross-section. Another possible extension of the work would be the bombardment of active DNA before the separation from protein.

Summary.—Pneumococcus transforming principle DNA has been bombarded with 2-m. e. v. electrons and 3.8-m. e. v. deuterons with indication by assay of activity remaining of a molecular weight of about 6,000,000.

Acknowledgments.—We were assisted in the cyclotron work by Mr John Preiss. We are indebted to Dr. A. O. Allen and to Dr. J. Saltick of Brookhaven for the use of the Van de Graaff machine and for assistance in those runs, and to Mr. Marvin Slater for the run with the linear accelerator.

* Assisted by the U. S. Atomic Energy Commission.

[1] Griffith, F., *J. Hyg., Camb.*, **27,** 113 (1928).
[2] Dawson, M. H., and Sia, R. H. P., *J. Exp. Med.*, **54,** 681 (1931).
[3] Avery, O. T., MacLeod, C. M., and McCarty, M., *Ibid.*, **79,** 137 (1944).
[4] Svedberg, T., and Brohult, S., *Nature*, **143,** 938 (1939).
[5] Lea, D., Smith, K. M., Holmes, B., and Markham, R., *Parasitology*, **36,** 110 (1944).
[6] Pollard, E., and Forro, F., *Arch. Biochem. Biophys.*, **32,** 256 (1951).
[7] Pollard, E., Buzzell, A., Jeffreys, C., and Forro, F., *Ibid.*, **33,** 9 (1951).
[8] McCarty, M., and Avery, O. T., *J. Exp. Med.*, **83,** 97 (1946).
[9] Taylor, H. E., *Ibid.*, **89,** 399 (1949).
[10] MacLeod, C. M., and Mirick, G. S., *J. Bact.*, **44,** 277 (1942).
[11] Davidson, J. M., *Biochemistry of Nucleic Acids*, Chap. IV, Wiley, 1950.
[12] Siri, W. E., *Isotopic Tracers and Nuclear Radiations*, Chap. III, McGraw-Hill, 1949.
[13] Gulland, J. M., Jordan, D. O., and Threlfall, C. J., *J. Chem. Soc.*, **1947,** 1129.

*STUDIES ON THE HEMOGLOBIN OF COOLEY'S ANEMIA AND COOLEY'S TRAIT**

BY ALEXANDER RICH†

GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated by Linus Pauling, January 26, 1952

The diseases sickle cell anemia and Cooley's anemia (also known as thalassemia or Mediterranean anemia) have associated with them "minor" or "trait" forms of the disease in which the symptomatology is minimal, and the abnormalities in the red cells are correspondingly less. Extensive genetic studies carried out on these diseases[1–4] have led to the conclusion that in each case the trait is inherited as a Mendalian dominant, the person

being heterozygous in the sickle cell gene or Cooley's gene. Homozygosity in either of these genes leads to the corresponding anemia.

Our understanding of sickle cell anemia has improved considerably with the discovery of an abnormal hemoglobin and a molecular interpretation of the pathological behavior of the red cell.[5] In that disease it was possible to demonstrate heterozygosity and homozygosity directly, by observing varying amounts of the abnormal hemoglobin.

Liquori[6] has recently reported the presence of fetal hemoglobin in cases of Cooley's anemia. We have carried out a number of experiments which support his findings, and have extended the investigation to include cases of Cooley's trait, for which no fetal hemoglobin was found. A hypothesis is advanced to explain these facts.

Methods.—Blood from eleven individuals, representing five different families, was investigated (table 1). All of the five patients with Cooley's

TABLE 1

PATIENT	AGE	SEX	RELATION	DIAGNOSIS
S. F.	18 mos	F	...	Cooley's anemia
D. F.	3 yrs	M	Brother of S. F.	Cooley's trait
M. F.	Adult	F	Mother of S. F.	Cooley's trait
V. P.	3 yrs	M	...	Cooley's anemia
J. P.	5 yrs	M	Brother of V. P.	Cooley's trait
J. C.	Adult	F	Maternal Aunt of V. P.	Cooley's trait
P. M.	Adult	F	Maternal Aunt of V. P.	Cooley's trait
M. L.	19 mos	F	...	Cooley's anemia
J. L.	Adult	M	Father of M. L.	Cooley's trait
B. S.	20 mos	F	...	Cooley's anemia
W. B.	18 mos	F	...	Cooley's anemia

anemia were typical in that they came from families of Italian extraction and displayed the characteristic hematological, roentgenographic and physical findings.[7]

None of the patients with Cooley's anemia had been transfused within three months. Patients B. S. and W. B. had never received transfusions; M. L. had been transfused 101 days before our blood specimen was taken, and the other two had not been transfused for over four months.

Blood was collected from these patients using sodium citrate as an anticoagulant, and stroma-free concentrated solutions of hemoglobin were prepared by the method of Drabkin.[8] Blood samples from new born infants were taken from the placental end of the cord and prepared by the same method. The carbon monoxide derivative was used for both the electrophoretic and spectrophotometric studies.

Electrophoresis.—Two types of electrophoresis experiments were undertaken. Those at 0.1 ionic strength (potassium phosphate buffer) and *p*H

range 5.9 to 7.0 were carried out in an attempt to separate components without introducing large boundary anomalies. Another group of experiments were performed in a 0.03 ionic strength sodium phosphate buffer at *p*H 8.2, in which the ratio of protein concentration to buffer ions is large and the resultant boundary anomalies permit the detection of components which differ only slightly in mobility.[9] The latter experiments were modifications of a procedure described fully by Beaven, Hoch and Holiday[10] in an investigation of adult and fetal hemoglobins.

The carbonmonoxyhemoglobin solutions were diluted with the phosphate buffer preparation to a concentration of 0.5 to 1.0%, and were then

TABLE 2

ELECTROPHORETIC MOBILITY AS A FUNCTION OF *p*H

PATIENT	DIAGNOSIS	*p*H OF BUFFER	MOBILITY $\times 10^5$ $CM^2\ VOLT^{-1}\ SEC^{-1}$
S. F.	Cooley's anemia	5.93	+1.88
	Cooley's anemia	6.9	−0.27
V. P.	Cooley's anemia	6.50	+0.45
		6.50	+0.50
		6.92	−0.14
		6.98	−0.29
D. F.	Cooley's trait	6.91	−0.33
M. F.	Cooley's trait	5.93	1.81
		6.9	−0.20
		6.91	−0.19
J. P.	Cooley's trait	5.93	1.84
		6.88	−0.20
		6.91	−0.19
J. C.	Cooley's trait	5.93	1.78
		6.75	0.04
		6.88	−0.19
		6.90	−0.16

[a] Average of ascending and descending limbs.

Buffer = 0.1 ionic strength phosphate; gradient = 3.8–6.7 volts cm^{-1}; protein concentration = 0.8–1.0 g per 100 ml.

dialyzed against the buffer at 4°C. for 12 to 24 hours in a carbon monoxide atmosphere. Values of *p*H were read on samples of the dialyzate after they had come to room temperature.

The modified Tiselius apparatus described by Swingle[11] was used. The experiments at 0.1 ionic strength employed potential gradients of 3.8 to 6.7 volts per cm. and those at 0.03 ionic strength employed gradients of 17.5 to 18.7 volts per cm. In the latter experiments, the boundaries were kept within the cell compartment by counter compensation. Before each electrophoresis the diluted carbonmonoxyhemoglobin solution was run in the preparative ultracentrifuge at 40,000 g for 40 minutes. It was found that this procedure effectively removed small amounts of colorless (presumably stromal) contaminants.

Alkaline Denaturation.—Some experiments were performed to determine the resistance of the hemoglobin from Cooley's anemia patients to alkaline denaturation. A small volume (0.05 ml) of the blood sample was diluted with 3 ml of N/12 NaOH to produce a 0.2% carbonmonoxyhemoglobin solution. This was read spectrophotometrically at 650 mμ, as suggested by Jonxis,[12] and the formation of the brown alkaline hemochromogen was recorded as a function of time.

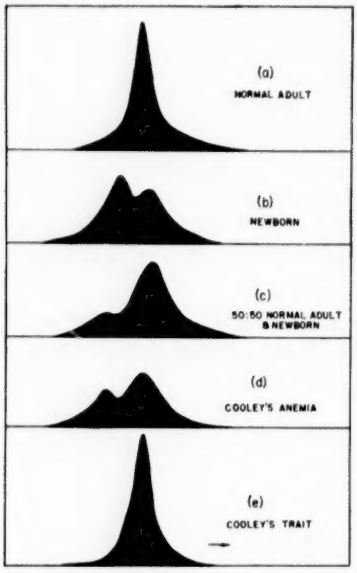

FIGURE 1

Electrophoretic patterns of the ascending boundary of carbonmonoxyhemoglobins. (*a*) Normal adult hemoglobin. (*b*) Newborn cord blood. (*c*) 50:50 mixture of (*a*) and (*b*). (*d*) Cooley's anemia hemoglobin, patient M. L. (*e*) Cooley's trait hemoglobin, patient J. L. Sodium phosphate buffer at pH 8.2; ionic strength 0.03; carbonmonoxyhemoglobin concentration 1.08 g per 100 ml.; time 280 to 300 min. The boundary is moving to the right.

Spectrophotometry.—The concentrated carbonmonoxyhemoglobin solutions were diluted with carbon monoxide-saturated distilled water and were then read immediately in a Beckman spectrophotometer, model DU. The slit width was 0.25 mm, which corresponds to a spectral width at half maximum of 0.88 mμ. The dilutions were always made to give a standard optical density, 0.430 at 290 mμ.

For the determination of small differences in optical density, multiple readings were made and the standard error was determined.

Results.—In table 2 are listed the mobilities of some of the hemoglobin preparations at different *p*H values. Special emphasis was placed upon examination of the hemoglobin of individuals with Cooley's trait, since it was felt that these might represent cases with mixtures of different types of molecules, in analogy with sickle cell trait. However, only one peak was consistently evident. The mobility of the trait blood between the *p*H values 6.88 and 6.92 was -0.21 ± 0.05 cm^2 volt^{-1} sec^{-1}, and the average value for the two anemia samples also was -0.21. A comparison of these values with the mobility-*p*H results cited for normal carbonmonoxyhemoglobin[5] shows that they are equivalent within experimental error.

Because of Liquori's report,[6] mobilities were measured of mixtures of

hemoglobin from Cooley's anemia patients with hemoglobin from normal adults and newborns. In no case could two peaks be resolved, and the mobility results were similar to those cited above. Fetal hemoglobin was found to migrate effectively with the same mobility as the adult form in the *p*H range at 0.1 ionic strength.

Additional experiments were then carried out at the higher *p*H values and lower ionic strengths in which it has been possible to resolve fetal and adult hemoglobins electrophoretically.[10]

In figure 1 are reproduced some of the patterns obtained from various blood samples. It can be seen that normal adult carbonmonoxyhemoglobin (*a*) produces a nearly symmetric, rapidly moving peak, in contrast to cord blood (*b*), for which two peaks are observed. The pattern for a mixture of the two (*c*) permits identification of the faster component as adult hemoglobin and of the slower as fetal hemoglobin.

TABLE 3

ELECTROPHORETIC ANALYSES OF COOLEY'S ANEMIA AND FETAL CARBONMONOXYHEMOGLOBINS

BLOOD	APPARENT % OF FASTER COMPONENT	CORRECTED % OF FASTER COMPONENT
1. B. S.—Cooley's anemia	6.5	2.8
2. W. B.—Cooley's anemia	5.6	2.0
3. M. L.—Cooley's anemia	63.4	42.
4. Newborn	41.0	18.
5. Newborn and normal adult (50:50 mixture)	83.7	62.

Ascending boundary only; *p*H 8.2, ionic strength 0.1, concentration 1.08 g per 100 ml, time 280–300 min.

Hemoglobin from Cooley's anemia patients (*d*) produces a two-peak pattern, with mobilities equal to those for a newborn and adult mixture, while the hemoglobin from Cooley's trait (*e*) produces one peak, indistinguishable from the normal adult (*a*).

This analysis has been carried out on the hemoglobin from three cases of Cooley's anemia (table 3) and three cases of Cooley's trait. Each of the anemia patterns shows the two components described above, while the trait patterns show only one. These results for Cooley's anemia substantiate the observations of Liquori.

In a buffer medium of low ionic strength in which there are two components with only slightly differing mobility, the faster component tends to concentrate itself ahead of the boundary of the slower component to balance the decrease in migration velocity. By using Svensson's equation[13] describing the conductivity change across a boundary, Hoch-Ligeti and Hoch[14, 9] have derived an expression relating the areas in the electrophoretic pattern of a two-component system to the concentrations. They assume

absence of interaction between the two components and linear dependence of migration velocity on concentration. These approximations have been justified for the hemoglobin system in 0.03 ionic strength.[9] We have utilized their equation to calculate the amounts of fetal and adult hemoglobins in the cases of Cooley's anemia. The apparent and calculated percentages of the faster moving (presumably normal adult) component are given in table 3.

Patients B. S. and W. B. have a very small amount of the faster moving material. It is interesting to note that the specimens B. S. and W. B. represent unusually pure samples of fetal hemoglobin. Newborn infants have 15 to 20% of adult hemoglobin, as observed electrophoretically, while even the 20-week fetus is found to have 6%.[10] This nearly pure preparation of fetal hemoglobin can be used for the calibration of spectrophotometric readings.

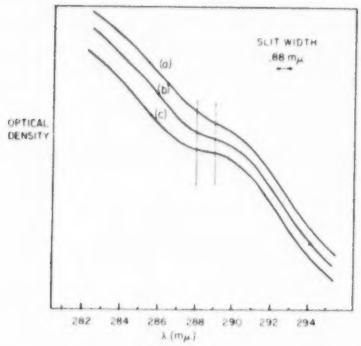

FIGURE 2

Ultra-violet absorption spectrum of carbon: monoxyhemoglobins. (*a*) Adult; (*b*) 50:50 adult-newborn mixtures; (*c*) newborn. The variation in the difference in optical density between 288 and 289 mμ can be readily seen. The curves actually all go through the same point at 290 mμ but are displaced vertically here for clarity.

Alkaline Denaturation.—We have carried out experiments on alkaline denaturation, with results in agreement with those of Liquori. Cooley's anemia hemoglobin was found to denature more slowly than normal adult hemoglobin; samples W. B. and B. S., with only 2 to 3% of the faster component, denature more slowly than samples of hemoglobin from newborns, in comformity with the electrophoretic analyses.

Spectrophotometry.—Jope[15] discovered a difference in the ultraviolet absorption spectra of fetal and adult hemoglobins. The inflection point of the tryptophan fine structure band is found at 2910 A in fetal and at 2898 A in normal adult hemoglobin. Beaven, Holiday and Jope[16] have measured the position of the inflection point in a series of fetal and infant bloods and have correlated it with the ratio of fetal and adult hemoglobin. Beaven, Hoch and Holiday[10] have shown that the position of the inflection point for artificial fetal-adult mixtures is not a linear function of concentration, since a small amount of fetal hemoglobin has a marked effect in producing a shift in the inflection point. Thus, they find their spectral method only of limited usefulness.

In the course of checking their work, we have devised a method which

we believe is of somewhat greater utility both in instrumentation and range of applicability. Curves such as those illustrated in figure 2 were obtained on the Beckman Model DU spectrophotometer. Curve (*a*) is the absorption spectrum of normal adult carbonmonoxyhemoglobin when examined with a slit width of 0.88 mμ, curve (*c*) is the spectrum for a sample of newborn carbonmonoxyhemoglobin, and curve (*b*) is the spectrum of a 50:50 mixture of the adult and newborn hemoglobins, the total concentration being constant. The increased prominence of the tryptophan shoulder can be seen in curve (*c*), although there is no resolution of a separate peak. Instead of attempting to determine the position of the inflection point, we measure the change in optical density for the interval 288 to 289 mμ.

If Beer's law is valid, the difference in optical density over this interval should be a linear function of the concentration of fetal or adult hemoglobin.

TABLE 4

SPECTROPHOTOMETRIC AND ELECTROPHORETIC ANALYSES OF COOLEY'S ANEMIA AND COOLEY'S TRAIT CARBONMONOXYHEMOGLOBIN

BLOOD	$\Delta 288\text{–}289 \times 10^3$	% ADULT HEMOGLOBIN	
		SPECTROPHOTOMETRIC	ELECTROPHORETIC
1. B.S.—Cooley's anemia	0	2.5	2.8
2. W. B.—Cooley's anemia	0.5	7.5	2.0
3. M. L.—Cooley's anemia	4.5	35	42
4. J. L.—Cooley's trait	12.5	95	100
5. J. C.—Cooley's trait	13.0	101	100

To test the assumption of Beer's law, artificial mixtures of adult and newborn hemoglobins were prepared, and the differences in optical density between 288 and 289 mμ ($\Delta 288\text{–}289$) were measured. All of these mixtures were prepared to a constant total concentration initially, and the percentage of adult hemoglobin in the newborn sample used was determined electrophoretically as cited above. The results, plotted in figure 3, show a linear dependence of $\Delta 288\text{–}289$ upon the concentration of adult hemoglobin.

Determination of small differences in optical density is possible provided one exercises great care in the measurements. By taking multiple measurements in the region near 0.400, optical densities can be determined to ± 0.0005. This corresponds to an error of $\pm 6\%$ in the determination of the amount of the adult component.

The accuracy of the curve in figure 3 was tested by determining $\Delta 288\text{–}289$ for various samples of hemoglobin, the fetal-adult compositions of which had been determined electrophoretically. The percentages of adult hemo-

globin as obtained by the two methods are given in table 4. They are found to be in agreement to within 7%.

Discussion.—Hoch[9] has described a faster moving component present in normal adult hemoglobin in concentrations of 2 to 3%. Our cases B. S. and W. B. have 2 to 3% of a component which migrates faster than fetal hemoglobin and could be either normal adult hemoglobin or the small component described by Hoch. The mobility data are not sufficiently accurate to settle this question.

As B. S. and W. B. had never received transfusions, they have continued to live beyond the fetal stage with use of fetal hemoglobin alone as oxygen

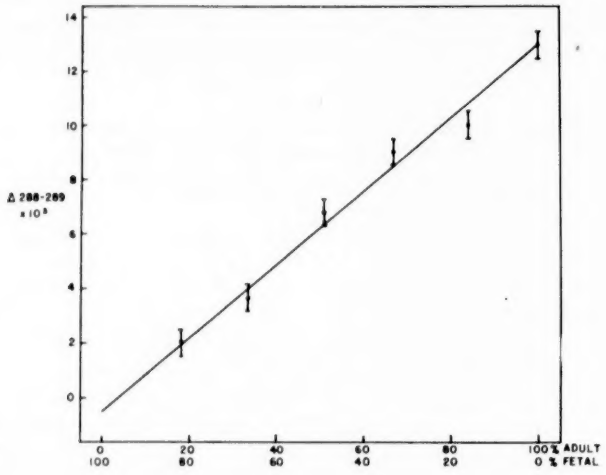

FIGURE 3

The change in the difference in optical density between 288 and 289 mμ as a function of adult hemoglobin concentration in fetal-adult mixtures.

carrier. To our knowledge, these are the first such cases reported in the literature.

Because M. L. had received a transfusion 101 days before the present sample was taken, we do not know with certainty what percentage of her normal adult hemoglobin was derived endogenously. Normal adult erythrocytes continue to live out their normal life span when transfused into patients with Cooley's anemia.[17] From a record of M. L.'s transfusion and hemoglobin levels, we estimate that one-half or less of her adult hemoglobin was obtained by transfusion. Liquori estimates that his two eight-year-old patients have 40 to 60% fetal hemoglobin.

We must explain why the homozygous patient with Cooley's anemia has large amounts of what we believe to be fetal hemoglobin, nearly 100%, while the heterozygous parent or sibling has no fetal hemoglobin.

Sickle cell disease differs from Cooley's anemia in that the presence of one mutant allele in the heterozygote results in the production of an altered hemoglobin molecule. Further, the homozygous patient with Cooley's anemia produces a molecule which appears not to be unique to the disease, but to be identical with a hemoglobin molecule normally produced at an earlier stage of development. The disease has prevented the cessation or has reactivated a metabolic process which normally stops at birth.

In order to bring this about, the Cooley's gene has both effectively to block the production of normal hemoglobin and to cause the continued production of fetal hemoglobin. We suggest that the primary effect of Cooley's gene is in some way to prevent the production of normal erythrocytes, and that there then occurs the continued production of erythrocytes containing fetal hemoglobin as a result of the consequent anemia. According to this interpretation fetal hemoglobin is not similar to sickle cell hemoglobin in being the aberrant product of an altered gene, but is instead the result of a compensatory response.

If this interpretation is valid, there should be other cases of severe anemia with which is associated the presence of fetal hemoglobin beyond the fetal stage. That this may be the case is suggested by the studies of Singer, Chernoff and Singer[18] in which they interpret a prolonged alkaline denaturation time as possibly indicating small amounts of fetal hemoglobin in various anemias.

The patients with Cooley's trait have one normal allele and one Cooley's allele, which together have the capacity to produce malformed, but stable erythrocytes. These patients have no severe anemia, and hence no strong hematopoetic stimulus, and it is observed that they do not produce cells containing fetal hemoglobin.

The exact nature of the defect in Cooley's anemia remains uncertain. The mutant gene may prevent some necessary step in the production of adult hemoglobin, or it may prevent the synthesis of normal adult stromal material, or in some other way impair normal red cell production. Any one of these defects would produce a severe anemia which, according to our postulate, would activate the production of cells containing fetal hemoglobin.

Acknowledgments.—The author wishes to express his thanks to Professor Linus Pauling and Dr. Harvey Itano for assistance in various parts of this investigation.

* Contribution No. 1655 from the Gates and Prellin Laboratories of Chemistry.

† Fellow in the Medical Sciences of the National Research Council.

[1] Neel, J. V., *Science*, **110**, 64 (1949).

[2] Neel, J. V., *Blood*, **6**, 389 (1951).
[3] Valentine, W. N., and Neel, J. V., *Arch. Int. Med.*, **74**, 185 (1944).
[4] Neel, J. V., and Valentine, W. N., *Genetics*, **32**, 38 (1947).
[5] Pauling, L., Itano, H. A., Singer, S. J., and Wells, I. C., *Science*, **110**, 543 (1948).
[6] Liquori, A. M., *Nature*, **167**, 950 (1951).
[7] We wish to express our thanks to Dr. Phillip Sturgeon of the Children's Hospital, Los Angeles, Calif., and Dr. Wayne Borges of the Children's Medical Center, Boston, Mass., for their cooperation in supplying blood samples and clinical information.
[8] Drabkin, D. L., *J. Biol. Chem.*, **164**, 703 (1946).
[9] Hoch, H., *Biochem. J.*, **46**, 199 (1950).
[10] Beaven, G. H., Hoch, H., and Holiday, E. R., *Ibid.*, **49**, 374 (1951).
[11] Swingle, S. M., *Rev. Sci. Instruments*, **18**, 128 (1947).
[12] Jonxis, J. H. P., *Haemoglobin*, Butterworth's Scientific Publications, London, 1949, p. 261.
[13] Svensson, H., *Arkiv Kemi Mineral. Geol.*, **22**, No. 10 (1946).
[14] Hoch-Ligeti, C., and Hoch, H., *Biochem. J.*, **43**, 556 (1948).
[15] Jope, E. M., *Haemoglobin*, Butterworth's Scientific Publications, London, 1949, p. 205.
[16] Beaven, G. H., Holiday, E. R., and Jope, E. M., *Discuss. Faraday Soc.*, **9**, 406 (1950).
[17] Kaplan, E., and Zuelzer, W. W., *J. Lab. Clin. Med.*, **36**, 517 (1950).
[18] Singer, K., Chernoff, A. I., and Singer, L., *Blood*, **6**, 413, 429 (1951).

THREONINE INHIBITION IN A STRAIN OF NEUROSPORA

BY C. O. DOUDNEY AND R. P. WAGNER

GENETICS LABORATORY OF THE DEPARTMENT OF ZOOLOGY, UNIVERSITY OF TEXAS

Communicated by J. T. Patterson, January 4, 1952

It is becoming increasingly apparent that metabolites normally present within the cell may act as inhibitors of some phases of metabolism while at the same time being necessary as intermediates in others.[1] This fact is of interest not only in the analysis of growth factor requirements, but also in the analysis of genetic blocks which prevent the production of cell metabolites. For it would follow that the inherited inability of an organism to synthesize a compound may not necessarily be due at all times to the absence of an enzyme catalyzing its synthesis, but rather to an inhibition of its synthesis by some other compound produced by the organism or supplied from the outside.

Work previously done in this laboratory on the pantothenicless mutants of *Neurospora crassa* has indicated that some such explanation as given above may apply to certain biochemical mutants of Neurospora, since the enzymes involved in the synthesis of pantothenic acid are present both in the pantothenicless mutants and the wild type strain.[2-4] Although the nature of the mechanism causing the enzyme system to be inactive in the

pantothenicless mutants is unknown, the presence of some type of inhibition is definitely indicated.

As a preliminary approach to the investigation of inhibitors as one of the factors involved in the formation of genetic blocks, a search was made for mutants of Neurospora which would grow like wild type on minimal medium, but be inhibited unlike wild type by specific compounds known to participate in the metabolism of Neurospora. One of the first of such mutants to be found, a strain which is strongly inhibited by threonine, is described in this communication together with experiments demonstrating the relief of threonine inhibition by other compounds.

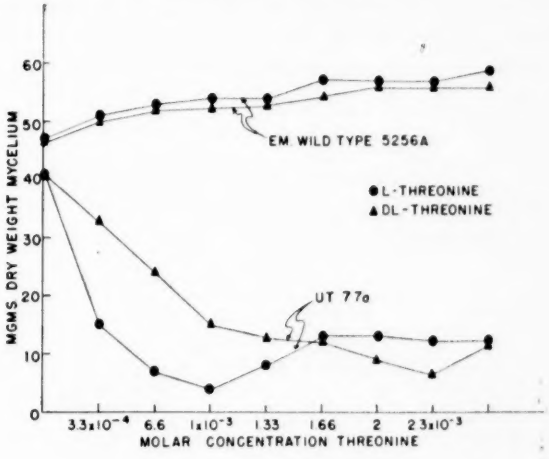

FIGURE 1

The effect of threonine on the growth of UT 77 and Wild Type, Emerson 5256.

Experimental.—The threonine inhibited strain, UT 77 originated from a single ascospore derived from a cross involving conidia irradiated with ultra-violet light. The method of Lein, Mitchell and Houlahan[5] for selective screening of biochemical mutants was used with one modification to select for the mutant ascospores. A high concentration of threonine and other amino acids was added to the agar plates upon which the spores were collected for germination. The threonine inhibited strain was separated from the other biochemical mutants detected by further screening in minimal agar tubes.

Compared to the wild type Emerson 5256 from which it was derived, the mutant UT 77 grows at about the same rate on minimal medium at 35°C., but about 50% slower than wild type at 25°C. Threonine inhibition

is most strongly expressed at 35°C. and at pH values below 5.8. At pH values above 5.8 threonine inhibition is slight and of short duration. Accordingly, in the experiments described below the temperature was maintained at 35°C. and the initial pH was adjusted to 4.8. Experiments were carried out using 125-ml. flasks to which 25 ml. of medium was added. All experimental data given are based on the mycelial dry weights for three flasks with identical medium. The minimal medium used was that described by Beadle and Tatum.[6]

Figure 1 compares the response of UT 77 and wild type after 60 hours growth on minimal medium supplemented with various concentrations of *L*- and *DL*-threonine. Maximum inhibition by *L*-threonine is given at a concentration of 1×10^{-3} molar, whereas 2.3×10^{-3} molar *DL*-threonine is required to give maximum inhibition. At these concentrations

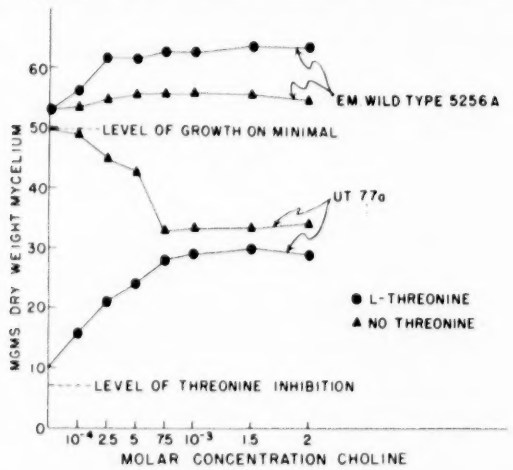

FIGURE 2

The inhibition of UT 77 by choline, and the relief of threonine inhibition by choline.

of threonine growth of UT 77 is only 10 to 20% of that produced on minimal medium. Threonine is, on the other hand, slightly stimulatory to the wild type strain.

Threonine inhibition of UT 77 is effective for a period of about 80 hours after which the mutant grows at a slowed rate compared to wild type. A final level of growth is attained in 152 hours (99 mg. dry weight of mycelium in 25 ml. of medium) which is equal to the final level of growth attained in 88 hours by the mutant on a medium free of added threonine. The final level of growth of UT 77 on minimal surpasses that of the wild type by 35%.

In addition to threonine, choline also inhibits the growth of UT 77. Choline at concentrations of 7.5×10^{-4} molar and higher inhibits growth to 65% of that obtained on minimal medium in 60 hours (Fig. 2). Choline does not inhibit the wild type strain. Complete relief of choline inhibition can be obtained by the addition of as little as 0.15 μg. of thiamin per ml. of medium, (table 1). No other B-vitamin tested shows this relief.

TABLE 1

THIAMIN RELIEF OF CHOLINE INHIBITION

CHOLINE CONCENTRATION	DRY WEIGHT MYCELIUM, MG.			
	UT 77		WILD TYPE	
	NO THIAMIN	THIAMIN[a]	NO THIAMIN	THIAMIN[a]
0	40	41	42	44
7.5×10^{-4} *M*	28	46	48	52
1.0×10^{-3} *M*	27	45	47	48
1.5×10^{-3} *M*	28	49	52	50

[a] Thiamin concentration 0.15 μg. per ml.

Choline partially antagonizes the inhibition produced by threonine (Fig. 2). The maximum relief of threonine inhibition obtained gives about the same level of growth as obtained with choline alone.

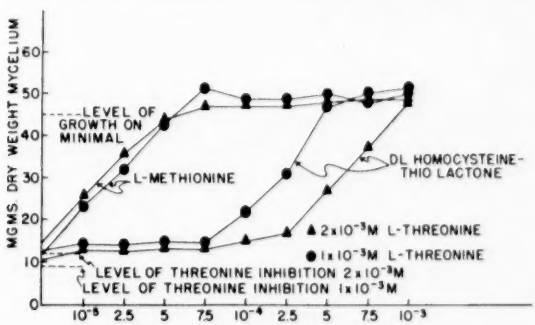

FIGURE 3

The relief of threonine inhibition of UT 77 by methionine and homocysteine-thiolactone.

Four compounds reverse threonine inhibition completely: *L*- or *DL*-methionine, *DL*-homocysteine, *DL*-homocysteine-thiolactone and *DL*-homoserine (Fig. 3 and table 2). *DL*-methionine is as effective as *L*-methionine and in both cases the reversal of threonine inhibition is non-competitive, since no increase of methionine concentration is required to effect the same level of relief of inhibition when the threonine level is increased. *DL*-homocysteine and *DL*-homoserine are appreciably less effective than methionine for the reversal of threonine inhibition (table 2).

Relief of threonine inhibition by homocysteine-thiolactone is clearly of a competitive type, since an equivalent increase in the concentration of the thiolactone is required to antagonize an increased amount of threonine in the medium (Fig. 3). Homoserine, while slightly more effective than homocysteine in the competitive relief of threonine inhibition in the lower concentration range, requires a many fold increase in concentration to relieve threonine inhibition at higher concentrations of threonine (table 2). It should be noted in this respect that homocysteine is a highly unstable compound in the culture medium, being subject to oxidation to homocysteine, which is inactive in the support of growth of Neurospora.[8]

TABLE 2

MINIMAL MOLAR CONCENTRATIONS OF VARIOUS COMPOUNDS REQUIRED TO COMPLETELY REVERSE L-THREONINE INHIBITION

	MOLAR CONCENTRATION L-THREONINE	
	1.0×10^{-2}	2.0×10^{-2}
L-methionine	5.0×10^{-5}	5.0×10^{-5}
DL-homocysteine	2.4×10^{-4}	4.9×10^{-4}
DL-homoserine	1.2×10^{-4}	7.5×10^{-3}
DL-homocysteine-thiolactone	5.0×10^{-4}	1.0×10^{-3}
Sulfanilamide	7.5×10^{-5}	7.5×10^{-5}

Sulfanilamide relieves the inhibition of threonine completely at 7.5×10^{-5} molar (Fig. 4). This relief is non-competitive (table 2). If the

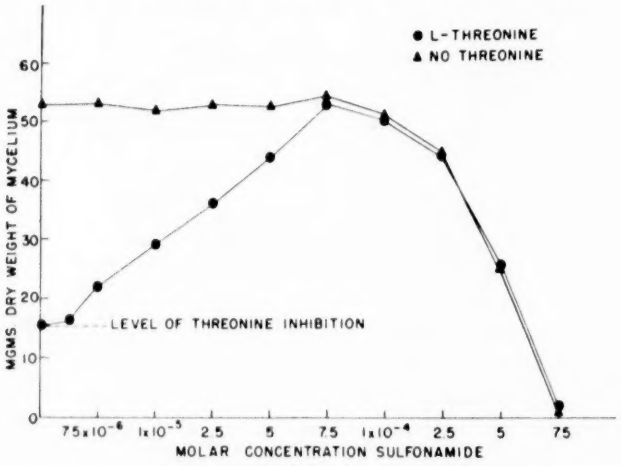

FIGURE 4

The relief of threonine inhibition of UT 77 by the sulfonamide, sulfanilamide.

sulfanilamide concentration is increased above 7.5×10^{-5} molar, growth of both UT 77 and the wild type is suppressed. As expected, *p*-amino-

benzoic acid antagonizes sulfanilamide reversal of threonine inhibition competitively; one mole of *p*-aminobenzoic acid antagonizes 100 moles of sulfanilamide. In the absence of sulfanilamide, *p*-aminobenzoic acid has no effect on the growth of the mutant even at concentrations as high as 1×10^{-5} molar.

Genetic studies of a cross of UT 77 to wild type show that threonine inhibition is due to a single gene mutation The difference between the inhibited and non-inhibited segregants obtained from the cross to wild type is decisive. For example, of 31 segregants derived from ascospores picked at random, 17 showed definite inhibition by threonine ranging from 55 to 100% with a mean of 77 ± 12%. On the other hand, the remaining 14 segregants were inhibited from 0 to 10% with a mean of 3.6 ± 2.3% inhibition. It may be assumed therefore that the threonine inhibition character is a distinct one dependent upon a single gene change.

Discussion.—The results of the growth response experiments demonstrate that UT 77 grows as well as wild type controls at 35°C. on minimal medium

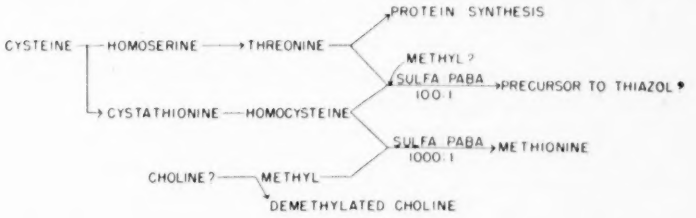

FIGURE 5

The relationship postulated for threonine, homocysteine and threonine metabolism in Neurospora.

in flask cultures, but is markedly inhibited by the addition of the amino acid, threonine to the medium. *DL*-threonine is about one-half as active as *L*-threonine, suggesting that the *D*-isomer is inactive in the process and that the observed suppression of growth is specifically produced by *L*-threonine. This is in agreement with the finding that the threonineless mutants of Neurospora cannot use *D*-threonine for growth.[7]

Threonine inhibition can be reversed by methionine non-competitively, and by homocysteine and homoserine in a competitive manner at higher concentrations. Horowitz[8] has shown that homocysteine is a precursor of methionine in Neurospora. (See Fig. 5.) The results reported here indicate that threonine causes inhibition by reacting with homocysteine (or a product to which it is readily converted) in an essentially irreversible reaction. The postulated reaction may be assumed to prevent growth by engaging the homocysteine synthesized by the mutant strain, thus preventing the formation of methionine. Therefore the strain becomes

"methionineless" or almost so when threonine is in the medium. The addition of methionine is able to relieve threonine inhibition by satisfying the threonine created methionine deficiency. Since homocysteine relieves the inhibition of threonine competitively, it, or a product derived from it, would seem to be involved both in reacting with threonine and in methionine production. When sufficient homocysteine is available in the medium to saturate the reaction with threonine, the synthesis of methionine can proceed at an adequate rate allowing usual growth.

Homoserine is believed to be a common precursor of both threonine and methionine, the latter being formed through cystathionine and homocysteine as intermediates.[7] Recently homoserine was demonstrated for the first time in a biological system when Fling and Horowitz[9] showed it to be present in extracts of a methionineless mutant of Neurospora. Table 2 shows that in UT 77 homoserine surpasses the effectiveness of homocysteine in the competitive relief of inhibition by low concentrations of threonine, while it becomes much less effective at higher concentrations of threonine in producing complete relief. Presumably, therefore homoserine exerts its effect by becoming converted to homocysteine at a rate sufficient to antagonize lower concentrations of threonine, but inadequate to effect complete relief of threonine inhibition when threonine is in the medium at higher concentrations.

The methyl group of methionine is thought to come from choline in Neurospora.[10-12] Mutant strains which cannot make their own choline can use methionine as a partial replacement for choline. The inhibition observed in the case of choline can be best explained by assuming that choline accelerates the production of methionine by donating methyl groups to homocysteine, thereby depriving the threonine-homocysteine reaction of homocysteine, and reducing the synthesis of an essential metabolite. This concept of the action of choline in the system is supported by the fact that choline partially antagonizes the inhibition by threonine indicating that the production of methionine from homocysteine is aided by the presence of appreciable amounts of choline in the medium.

It can be concluded that *p*-aminobenzoic acid has a role in the postulated threonine-homocysteine reaction system because of the sulfanilamide relief of threonine inhibition and its reversal by *p*-aminobenzoic acid. It must be assumed that the amount of sulfanilamide required to reverse threonine inhibition completely is sufficient to antagonize the *p*-aminobenzoic acid produced by the mutant strain to a level limiting to the postulated threonine-homocysteine reaction, and thereby making adequate amounts of homocysteine available for the production of methionine.

According to the hypothesis presented above choline acts to deprive the postulated reaction between homocysteine and threonine of homocysteine. The fact that thiamin in low concentrations will reverse choline

inhibition suggests that the reaction between homocysteine and threonine may be involved in the biosynthesis of thiamin. Thus thiamin added to the medium would supply the deficit produced by the action of choline and allow normal growth. Further experiments performed in this laboratory support the view that threonine is a possible precursor to the thiazole portion of thiamin. These experiments will be described in another publication.

The results reported here with UT 77 can best be explained by assuming that when either of the precursors threonine or choline of the competing reactions is available in excess amount, that reaction is caused to go at an excessive rate, thereby producing a deficiency of the common precursor, homocysteine. Since inhibition is not produced in the wild type strain by either choline or threonine, a mechanism must exist in wild type which exerts a controlling influence over the rates of these competing reactions, preventing either from going to excess and depriving the other of the common precursor. These hypotheses together with those based on previous work with methionineless and threonineless mutants of Neurospora[7-9, 15, 16] which establish the relationship of threonine and methionine in biosynthesis are summarized in fig. 5.

Emerson and Cushing[16] and Emerson[17] report a mutation, *sfo* in Neurospora which causes a requirement for sulfonamide at 35°C. and at pH values below 6. It was postulated by Zalokar[18] that this strain is inhibited by the *p*-aminobenzoic acid normally produced, and that sulfonamide promotes growth by antagonizing this inhibition. Zalokar,[19] using the *sfo* character in combination with a genetic block preventing the synthesis of methionine, showed that low concentrations of methionine support growth in the absence of sulfonamide while higher concentrations required that sulfonamide be present in order for growth to take place. A similar effect was found in the case of homocysteine. It was further found that threonine as the sole supplement to minimal medium will permit the *sfo* strain to grow. These results have been interpreted as meaning that the mutation responsible for the sulfonamide requirement controls a deleterious reaction which, using a precursor of methionine as a substrate, is catalyzed by *p*-aminobenzoic acid and interferes with the normal utilization of threonine.

It seems probable that the same reaction is responsible for the unusual growth responses described for the *sfo* mutant and UT 77. In each mutant the sulfonamide to *p*-aminobenzoic acid ratio demonstrated is approximately 100,[18] and in addition the mutant phenotype is best expressed at 35°C. and at the lower pH ranges. Furthermore, the interacting substrates are the same for each mutant, homocysteine or its immediate product, and threonine.

The differences observed between the *sfo* mutant and UT 77 must be

due to a specific genetic difference which allows the same reaction to exert a detrimental effect on growth, but in the case of the *sfo* strain interferes with an aspect of threonine utilization, while in UT 77 it prevents the formation of methionine when excessive amounts of threonine are available. It is to be noted that the *sfo* strain produces enough homocysteine to inhibit itself in minimal medium, but that UT 77 will grow about as well as the wild type strain in minimal at 35°C. Presumably, therefore, there is a balance in the production of the precursors to the postulated reaction in this case.

From a genetic viewpoint, the behavior of UT 77 supports the concept that the inability to mediate a chemical synthesis by an organism due to a genetic block is not necessarily due to the absence or inactivation of an enzyme required for that synthesis, but may be due to a metabolite inhibition resulting from the mutation.

The fact that inhibited strains such as UT 77 may be induced by the same means as are used to induce the ordinary biochemical mutants with nutritional deficiencies, makes it evident that the phenomenon of inhibition must be regarded as an important possible causative factor of genetic blocks. Although UT 77 is not a biochemical mutant on minimal medium, it is on exogenous threonine, and it is not too far fetched to assume that a more drastic mutation may cause it to be inhibited by its own threonine, and thus become a biochemical mutant on minimal medium requiring methionine or its precursors. The present data by no means prove that genetic blocks are produced in this way in biochemical mutants, but they do definitely indicate that further work predicated on this assumption is warranted.

Summary.—A new mutant strain of Neurospora is described which gives wild type growth on minimal medium at 35°C., but which is strongly inhibited by *L*-threonine. This inhibition is completely overcome by methionine, homocysteine, homoserine and sulfanilamide, and partially overcome by choline. Choline also inhibits the mutant, but to a lesser degree than threonine. The choline inhibition is relieved by thiamin. The significance of these observations is discussed, and their possible bearing on the understanding of the nature of genetic blocks considered.

[1] Snell, E. E., *Plant Growth Substances*, Univ. of Wisconsin Press, 431–446 (1951).
[2] Wagner, R. P., and Guirard, B. M., these PROCEEDINGS, **34,** 398–402 (1948).
[3] Wagner, R. P., *Ibid.*, **35,** 185–189 (1949).
[4] Wagner, R. P., and Haddox, C. H., *Am. Nat.*, **85,** 319–330 (1951).
[5] Lein, J., Mitchell, H. K., and Houlahan, M. B., these PROCEEDINGS, **34,** 435–442 (1948).
[6] Beadle, G. W., and Tatum, E. L., *Am. J. Bot.*, **32,** 678 (1945).
[7] Teas, H. J., Horowitz, N. H., and Fling, M., *J. Biol. Chem.*, **172,** 651–658 (1948).
[8] Horowitz, N. H., *Ibid.*, **171,** 255–264 (1947).
[9] Fling, M., and Horowitz, N. H., *Ibid.*, **190,** 277–285 (1951).

[10] Horowitz, N. H., and Beadle, G. W., *Ibid.*, **150,** 325–333 (1943).
[11] Horowitz, N. H., Bonner, D., and Houlahan, M. B., *Ibid.*, **159,** 145–151 (1945).
[12] Horowitz, N. H., *Ibid.*, **162,** 413–419 (1945).
[13] Shive, W., *J. Am. Chem. Soc.*, **69,** 725 (1947).
[14] Stetten, M. R., and Fox, C. L., *J. Biol. Chem.*, **161,** 333 (1945).
[15] Teas, H. J., Thesis, California Institute of Technology (1947).
[16] Emerson, S., and Cushing, J. E., *Federation Proc.*, **5,** 379–389 (1946).
[17] Emerson, S., *J. Bact.*, **54,** 195–207 (1947).
[18] Zalokar, M., these PROCEEDINGS, **34,** 32–36 (1948).
[19] Zalokar, M., *J. Bact.*, **60,** 191–203 (1950).

*OBSERVATIONS ON THE BEHAVIOR OF SUPPRESSORS IN NEUROSPORA**

By MARY B. MITCHELL AND HERSCHEL K. MITCHELL

KERCKHOFF LABORATORIES OF BIOLOGY, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated by G. W. Beadle, January 14, 1952

A suppressor of pyrimidineless *3a* (37301) and some aspects of the behavior of the suppressed mutant have been described earlier.[1] The observation that lysine, ornithine, citrulline and arginine influence growth responses of the suppressed mutant suggested studies of the behavior of recombinants involving *pyr 3a* and *s* and mutants having requirements for these amino acids. Effects of the pyrimidineless mutant and its suppressor upon certain lysine-requiring mutants have been reported.[2] The present paper deals with a somewhat greater variety of interactions observed between *pyr 3a* and *s* and mutants which utilize proline, ornithine, citrulline or arginine.[3] These interactions include suppression of two non-allelic prolineless mutants by the pyrimidineless suppressor and partial suppression of *pyr 3a* by three non-allelic ornithineless mutants.

Effects of the pyrimidineless suppressor on mutants of the proline-arginine series.—The suppressed pyrimidineless mutant, *pyr 3a-s* (37301-s) was first described[1] as being characterized on minimal medium by a lower growth rate than that of wild type, and by stimulation by pyrimidine, lysine or histidine. It has since been found that this is true only of some isolates. Others grow as rapidly as wild type, still others very nearly so, and the stimulations are very slight. The reported inhibition by ornithine, citrulline and arginine and its release by lysine remain properties of the recent isolates, however, although there may be differences in the effectiveness of the compounds as inhibitors. Proline is also inhibitory but very slightly so.

Crosses of *pyr 3a-s* were made to the following strains: 21863, prolineless; 35401, 44207 and 51506, which use proline, ornithine, citrulline or arginine;

27947, 29997 and 34105, which utilize ornithine, citrulline or arginine; 30300 and 33442, which utilize citrulline or arginine. For convenience these mutants have been renamed in the following fashion:

prolineless

21863 = *prol 1* 35401 = *prol 2* 44207 = *prol 3* 51506 = *prol 4*

ornithineless

27947 = *orn 1* 29997 = *orn 2* 34105 = *orn 3*

citrullineless

30300 = *cit 1* 33442 = *cit 2*

Growth responses and genetic relationships of these strains have been described by Srb and Horowitz,[3] Srb[4] and Srb, Fincham and Bonner,[5] who have shown that, except for *prol 3* and *4*, which behave as alleles, strains among these which use the same compounds are genetically distinct.

Ascus dissections were done on minimal agar plates as described by Haas *et al.*,[6] and the spores were heat treated and allowed to germinate on the plates. The germinated spores were transferred to tubes containing minimal agar medium supplemented with cytidine and proline or arginine and the resulting cultures classified by means of tests in tubes of liquid minimal medium with appropriate supplements. Members of spore pairs were isolated and cultured together as reported by Fincham[7] and Haskins and Mitchell.[8]

In tables 1 and 2 the types of asci observed are tabulated so as to show the phenotypes, but not the order in the ascus, of the spore pairs. The ornithineless and citrullineless mutants are designated as *arg* in order to simplify the tabulation in table 2. The suppressor, *s*, is represented only in those spore pairs in which its presence could be detected phenotypically by the qualitative classification tests used. From table 1 it may be seen that from the crosses *pyr 3a-s* × *prol 2*, *3* and *4* the prolineless mutant appeared in only one of the four spore pairs in asci of types 2, 5, 7 and 12, and in types 3, 6 and 10 it did not appear in any spore pair. Hence, it seems that *prol 2*, *3* and *4* are suppressed, and the types of segregation observed are consistent with the view that they are suppressed by *s*. This was confirmed by recovering the three prolineless mutants from out-crosses of phenotypically *pyr-s* strains derived from asci of types 2 and 3, and also by recovering them from out-crosses of phenotypically wild strains from asci of types 10 and 12. For the sake of clarity it should be pointed out that only one of the *pyr-s* spore pairs from a type 2 ascus, and only one + spore pair from a type 12 ascus carried the prolineless mutant, since it appeared, unsuppressed, in one spore pair in asci of these types.

There is no indication that *prol 1* is suppressed since it appeared in two spore pairs from each ascus, although in types 4, 9 and 11 it presumably

combined with *s* in at least one spore pair, as evidenced by the appearance of unsuppressed *pyr*. As table 2 shows, the same result was obtained with the ornithineless and citrullineless mutants since they appear to have combined with *s* in asci of types 2, 4, 5 and 6. The presence of *s* in an unsuppressed *arg* strain from each cross was demonstrated by crossing these to *pyr 3a* and recovering *pyr 3a-s*.

TABLE 1

CROSSES OF *pyr 3a-s* TO *prol 1, 2, 3* AND *4*

	TYPES OF SEGREGATION				*prol 1*	*prol 2*	*prol 3*	*prol 4*
1	*pyr-s*	*pyr-s*	*prol*	*prol*	1	3	..	5
2	*pyr-s*	*pyr-s*	*prol*	+	..	2	1	4
3	*pyr-s*	*pyr-s*	+	+	..	1	1	4
4	*pyr-s*	*pyr*	*prol*	*prol*	1	..	..	..
5	*pyr-s*	*pyr*	*prol*	+	..	3	2	..
6	*pyr-s*	*pyr*	+	+	..	3	..	1
7	*pyr-s*	*pyr-prol*	+	+	..	..	4	..
8	*pyr-s*	*pyr-prol*	*prol*	+	2	1	3	..
9	*pyr*	*pyr*	*prol*	*prol*	1	..	..	..
10	*pyr*	*pyr*	+	+	..	2	..	2
11	*pyr*	*pyr-prol*	*prol*	+	2	..	..	..
12	*pyr*	*pyr-prol*	+	+	..	3	5	5
13	*pyr-prol*	*pyr-prol*	+	+	1	6	1	2
					8	24	17	23

NOTE: Only the phenotypes of spore pairs are shown, not the complete genotypes nor the order in the ascus.

TABLE 2

CROSSES OF *pyr 3a-s* TO *orn 1, 2* AND *3* AND *cit 1* AND *2*

	TYPES OF SEGREGATION				*orn 1*	*orn 2*	*orn 3*	*cit 1*	*cit*
1	*pyr-s*	*pyr-s*	*arg*	*arg*	7	..	..	4	5
2	*pyr-s*	*pyr*	*arg*	*arg*	1	..	1	1	2
3	*pyr-s*	*pyr-arg*	*arg*	+	1	3	1	5	..
4	*pyr*	*pyr*	*arg*	*arg*	..	..	..	2	3
5	*pyr*	...	*arg*	*arg*	..	..	..	2	1
6	*pyr*	*pyr-arg*	*arg*	+	..	5	2	..	..
7	*pyr-arg*	*pyr-arg*	+	+	4	2	2	2	..
					13	10	6	16	11

NOTE: Only the phenotypes of spore pairs are given, not the complete genotypes nor the order in the ascus.

Growth responses of recombinants from the crosses were further tested by measuring dry weights of mycelium from cultures in 125-ml. flasks containing 20 ml. of Fries medium[9] with appropriate supplements. The suppressed prolineless mutants *prol 2-s*, *prol 3-s* and *prol 4-s* are not distinguishable from wild type on minimal medium and are not affected by

added proline or arginine. The suppressed double mutants *prol 2-pyr 3a-s* and *prol 3-pyr 3a-s* were not found to differ from *pyr 3a-s*.

Tests of recombinants of *s* with the ornithineless and citrullineless mutants have shown that *s* affects these in a way which may be considered the opposite of suppression. Three of these mutants, *orn 3*, *cit 1* and *cit 2*, grow slowly on unsupplemented medium, and in about 10 days reach a dry weight which approaches that of wild type after a 4-day growth period. As would be expected, they have lower requirements than do *orn 1* and *2*, which do not grow on unsupplemented medium. All of the five combinations with *s*, however, fail to grow a weighable amount on unsupplemented

TABLE 3

REPRESENTATIVE DRY WEIGHTS (IN MG.) FROM 125-ML. FLASK CULTURES AT 25°C. AFTER 4 OR 10 DAYS' GROWTH WITH THE FOLLOWING SUPPLEMENTS IN 20 ML. OF MEDIUM: ORNITHINE = DL ORNITHINE HCl, 1 MG.; CITRULLINE = DL CITRULLINE, 1 MG.; ARGININE = L ARGININE HCl, 1 MG.; CASEIN = ENZYME-HYDROLIZED, 40 MG.

	0		ORNITHINE		CITRULLINE		ARGININE		CASEIN
DAYS	4	10	4	10	4	10	4	10	4
orn 1	0	0	12	..	21.5	..	30.5	..	57
orn 2	0	0	9	..	27	..	29.5	..	..
orn 3	2	65	53	..	56	..	51	..	61
cit 1	3	40	2.5	50	63	..	46	..	58
cit 2	5.5	66	4	61	57	..	42.5	..	68.5
orn 1-s	0	0	0	tr.	13.5	..	21	..	39
orn 2-s	0	0	tr.	tr.	14	..	28	..	..
orn 3-s	0	tr.	tr.	tr.	20	..	33	..	40.5
cit 1-s	0	0	0	0	15.5	..	23	..	42
cit 2-s	0	tr.	0	tr.	19.5	..	31	..	47
pyr 3a	0	0	0	0	0	0	0	0	0
pyr 3a-orn 1	0	0	tr.	13.5	1.5	3.5	tr.	23.5	6
pyr 3a-orn 2	0	0	tr.	9.5	tr.	2	tr.	18	..
pyr 3a-orn 3	2.5	105.5	tr.	49.5	4.5	39	tr.	44.5	26.5
pyr 3a-cit 1	0	0	0	0	0	0	0	0	0

NOTE: tr. (trace) represents a quantity of mycelium which is visible but too small to be removed and weighed.

medium in 10 days, and give about the same response to citrulline or arginine as *orn 1* and *2*. Furthermore, the recombinants *orn 1-s*, *orn 2-s* and *orn 3-s* all fail to respond to ornithine even at a level which is 3 to 4 times higher than that required for half-maximum growth of *orn 1* and *2*. Hence the presence of *s* appears to convert these five mutants into strains with absolute requirements for citrulline or arginine. This is shown by data given in tables 3 and 4.

The effect of ornithineless mutants on the pyrimidine requirements of pyr 3a.—In classification tests of the progeny from the crosses considered in the preceding section the behavior of the double mutants of *pyr 3a* and *orn 1*, *2* and *3* suggested that they lacked the absolute pyrimidine require-

ment which is characteristic of *pyr 3a*. These double mutants were therefore obtained from crosses of the three ornithineless mutants to *pyr 3a* in order to make it certain that *s* was not present. Results from tests of these recombinants in flasks appear in tables 3 and 4. In *pyr 3a-orn 3* the pyrimidine requirement appears to be completely suppressed on unsupplemented medium, since this strain grows in the same manner as does *orn 3*. On medium supplemented with ornithine, citrulline or arginine growth is somewhat slower than on minimal medium, but more rapid growth was obtained with enzyme-hydrolyzed casein as supplement. *Pyr 3a-orn 1* and *pyr 3a-orn 2*, like *orn 1* and *2*, fail to grow on unsupplemented medium but grow slowly without pyrimidine when supplied ornithine or arginine, until they reach the dry weight characteristic of the

TABLE 4

REPRESENTATIVE DRY WEIGHTS (IN MG.) FROM 125-ML. FLASK CULTURES WITH 20 ML. OF MEDIUM, AFTER 4 DAYS' GROWTH WITH THE SUPPLEMENTS GIVEN BELOW

	0	1 MG.	2 MG.	4 MG.	8 MG.
	DL Ornithine Hydrochloride				
orn 1	0	13	24	44	80.5
orn 2	0	11	22	35.5	44.5
orn 3	2	57	54	59	65.5
orn 1-s	0	0	0	0	0
orn 2-s	0	tr.	tr.	tr.	tr.
orn 3-s	tr.	tr.	tr.	tr.	tr.
	Cytidine Sulfate—0.5 Mg. DL Citrulline				
pyr 3a-orn 1	0	3.5	6.5	18	73
pyr 3a-orn 2	0	4	49.5	77	65
pyr 3a-orn 3	3	64.4	89	82.5	63
pyr 3a-cit 1	0	2	2.5	1.5	1.5
pyr 3a	30.5	32	33	31	32

NOTE: tr. (trace) represents a quantity of mycelium which is visible but too small to be removed and weighed.

single ornithineless mutant on that level of the supplement. They grow even more slowly with citrulline as supplement, as would be expected for reasons which are given in the next section. After 14 days, however, with 4 mg. of DL citrulline per flask, the following dry weights were obtained: *pyr 3a-orn 1*, 109 mg.; *pyr 3a-orn 2*, 115 mg.

The difference in the rate of growth of *pyr 3a-orn 3* on minimal medium and on medium supplemented with ornithine, citrulline or arginine suggests that these compounds interfere with suppression of *pyr 3a* by *orn 3* as they do with suppression of *pyr 3a* by *s*. This is also suggested by the observation that growth of *pyr 3a-orn 3* is more rapid with lower than with higher levels of ornithine. The effect of lysine on the *pyr 3a-orn* double mutants has not been thoroughly investigated, but it appears to be com-

plicated by inhibition of the *orn* mutants by lysine. A significant stimulation of *pyr 3a-orn 3* by 6 mg. of L lysine hydrochloride was observed when this strain was cultured in the presence of 2 mg. of DL ornithine hydrochloride. With this ratio of lysine to ornithine *pyr 3a-orn 1*, *pyr 3a-orn 2*, *orn 1* and *orn 2* were completely inhibited whereas growth of *orn 3* was less affected. Thus it seems that lysine may exert the same effect on the *pyr 3a-orn* double mutants as it does on *pyr 3a-s* but that this is obscured by its inhibitory effect on the *orn* mutants.

It is considered unlikely that a change in the *pyr 3a* gene, such as back-mutation, is responsible for the apparent suppression of the pyrimidine requirement in these double mutants. The absence of an absolute requirement for pyrimidine was detected in *pyr-orn* double mutants among the progeny of six crosses involving two different isolates of both *pyr 3a* and *pyr 3a-s*. Segregants carrying *pyr 3a* alone, on the other hand, even those derived from asci which also produced double mutants, showed the absolute requirement. As a further check, the isolates of the double mutants which were used in the above experiments were crossed to wild type and *pyr 3a* recovered with its pyrimidine requirement unchanged.

The effect of pyr 3a on citrulline utilization.—As may be seen from data in table 4, the utilization of citrulline by *orn 1* and *2* and *cit 1* appears to be influenced by the presence of the *pyr 3a* gene. This effect is more marked in the case of *pyr 3a-cit 1*, which, in the presence of cytidine, gave a very meager response to citrulline at any level tested, whereas *pyr 3a-orn 1* and *pyr 3a-orn 2* appear to have merely increased requirements. There is no effect in the case of *pyr 3a-orn 3* nor has there been observed any effect upon ornithine utilization by *orn 1*, *2* and *3* nor upon arginine utilization by any of these strains. (*Pyr 3a-cit 2* has not yet been obtained because of rather close linkage which exists between these two loci.) Cytidine has not been found to affect the utilization of citrulline by the single *orn* and *cit* mutants, nor does citrulline interfere with cytidine utilization by *pyr 3a*.

It is of interest that, as table 4 shows, dry weights from the *pyr 3a-orn* double mutants cultured on cytidine plus citrulline were in some cases as much as twice that given by *pyr 3a* alone on the same amounts of cytidine and citrulline. This again demonstrates the partial suppression of the pyrimidine requirement in these double mutants.

Other recombinants.—There are known at present, in addition to *pyr 3a*, three non-allelic pyrimidineless mutants, *pyr 1* (263),[1, 10] *pyr 2* (38502),[1, 10] and *pyr 4* (36601).[11] It has already been reported that *pyr 1* and *2* are not suppressed by *s* and this has been found to be true also of *pyr* 4. These mutants, and *pyr 3a* as well, are inhibited by arginine when uracil is the only exogenous source of pyrimidine, but there is no inhibition if cytidine or uridine is added. No greater than half inhibition has been observed at any level of arginine tested, and there is no release by lysine. In double

mutants with *orn 3*, *pyr 1*, *2* and *4* were not suppressed and, in the presence of added cytidine, did not affect the behavior of *orn 3*.

The double mutants of *pyr 3a* with *prol* 2 and *3*, when tested in flasks, behaved as would be predicted from the properties of the single mutants. No effect of either mutant gene upon the requirement introduced by the other was observed. Some isolates of *pyr 3a-prol 1* behaved is this manner, but others, which grew on complete medium, failed to grow when supplied cytidine and proline. This has not been investigated further.

It has been reported[2] that 36703, a mutant which uses only arginine[3] showed no interaction with *pyr 3a* or *s*. More recently a histidineless mutant, C84[6] was tested, with the same result. In linkage tests[12] *pyr 3a* or one of its possible alleles, has been crossed to mutants requiring inositol, pyridoxin, pantothenic acid, choline, adenine, phenylalanine, isoleucine + valine and tryptophan, respectively, and no suppression of the pyrimidine requirement was observed.

SUMMARY OF SINGLE MUTANTS AND RECOMBINANTS

Single Mutants

prol 1	slight growth on minimal—uses proline.
prol 2, *prol 3*	grow slowly on minimal—use proline, ornithine, citrulline or arginine.
orn 1, *orn 2*	no growth on minimal—use ornithine, citrulline or arginine.
orn 3	grows slowly on minimal—uses ornithine, citrulline or arginine.
cit 1, *cit 2*	grow slowly on minimal—use citrulline or arginine.
pyr 1	slight growth on minimal—requires pyrimidine.
pyr 2	no growth on minimal at 35°C.—requires pyrimidine.
pyr 3a, *pyr 4*	no growth on minimal—require pyrimidine.
s	no phenotypic differences from wild type observed.

Recombinants

prol 1-s	no phenotypic differences from *prol 1* observed.
prol 2-s, *prol 3-s*	no phenotypic differences from wild type observed.
orn 1-s, *orn 2-s*, *orn 3-s*, *cit 1-s*, *cit 2-s*	no growth, or very slight growth on minimal—use citrulline or arginine but not ornithine.
pyr 1-s, *pyr 2-s*, *pyr 4-s*	no phenotypic differences from *pyr 1*, *2* and *4* observed.
pyr 3 a-s, *pyr 3a-prol 2-s*, *pyr 3a-prol 3-s*	growth on minimal approximates that of wild type—inhibition by proline, ornithine, citrulline or arginine relieved by lysine.

pyr 3a-prol 2, *pyr 3a-prol 3* — no growth on minimal—require pyrimidine + proline, ornithine, citrulline or arginine.

pyr 3a-orn 1, *pyr 3a-orn 2* — no growth on minimal—grow slowly without added pyrimidine on medium supplemented with ornithine, citrulline or arginine—in presence of cytidine the citrulline requirement is higher than that of *orn 1* or *2*.

pyr 3a-orn 3 — indistinguishable from *orn 3* on minimal—grows slowly without pyrimidine on medium supplemented with ornithine, citrulline or arginine—with cytidine present the response to these compounds is like that of *orn 3*.

pyr 3a-cit 1 — no growth on minimal or on medium supplemented with cytidine alone. In the presence of cytidine the response to arginine is like that of *cit 1-s*, but there is almost no response to citrulline.

pyr 1-orn 3 — slight growth on minimal—uses cytidine + ornithine, citrulline or arginine.

pyr 2-orn 3 — no growth on minimal at 35°C.—uses cytidine + ornithine, citrulline or arginine.

pyr 4-orn 3 — no growth on minimal—uses cytidine + ornithine, citrulline or arginine.

Discussion.—In order to simplify discussion of the observations presented, a summary of the various interactions involving *pyr 3a* and *s* is given below.

The Effects of the Suppressor of Pyrimidineless:

1. It suppresses *pyr 3a* in the absence of added proline, ornithine, citrulline or arginine, or, in the presence of these compounds if lysine also is added.[1]
2. It suppresses two genetically different mutants, *prol 2* and *3*, which require proline, ornithine, citrulline or arginine.
3. It suppresses the double mutants, *pyr 3a-prol 2* and *pyr 3a-prol 3*, which, in combination with it, respond to proline, ornithine, citrulline, arginine and lysine as does *pyr 3a-s*.
4. It prevents the slow growth characteristic of *orn 3*, *cit 1* and *cit 2* on unsupplemented medium.
5. It prevents the utilization of ornithine by the ornithineless mutants.
6. It interferes with the utilization of α-amino adipic acid by the lysineless mutant, *ly 1* (33933) unless arginine, citrulline or ornithine is added.[2]
7. In combination with *pyr 3a* it prevents utilization of α-amino adipic acid by *ly 1*.[2]
8. In combination with *pyr 3a* it interferes with the utilization of lysine by *ly 3* (4545) unless arginine is added.[2]

Effects of Pyrimidineless 3a:

1. It interferes with the utilization of citrulline by *orn 1* and *2* and almost completely stops citrulline utilization by *cit 1*.
2. It prevents the slow growth characteristic of *cit 1* on medium supplemented with cytidine.
3. It interferes with utilization of α-amino adipic acid by *ly 1* unless a small amount of lysine is added.[2]

Effect of orn 1 and 2 on pyr 3a:

They partially suppress *pyr 3a* on medium supplemented with ornithine, citrulline or arginine.

Effect of orn 3 on pyr 3a:

It suppresses *pyr 3a*, completely on minimal medium and partially on medium supplemented with ornithine, citrulline or arginine.

The pyrimidineless suppressor, then, acts as a suppressor of the recognized phenotypic effects of three genes, but when it is present in combination with any one of seven other mutant genes, it acts as an inhibitor, in the sense that it prevents strains carrying these genes from doing something which they are able to do if *s* is replaced by its "wild type" allele. Except for the difference in the degree of suppression, the "wild type" allele of *s* might be more justly called a suppressor. In any case there appears to be no reason for considering *s* to be in a different category from the other mutant genes involved in the interactions described here, since, for example, *pyr 3a* inhibits and the ornithineless mutants suppress. It is possible that "suppressors" which themselves introduce requirements are as frequent as those which do not, but their detection may be less likely. In cases reported by Zalokar[13] and Wagner and Haddox[14] suppression of one mutant by a second was predicted on the basis of growth responses of the first and this suppression was observed when the double mutant was prepared and tested. The double mutant, involving pantothenicless and tyrosine or phenylalanineless, described by Wagner and Haddox, is perhaps unique in that, although it does not grow normally in the absence of pantothenic acid, it is not stimulated by addition of this growth factor.

It now seems unlikely that *s* takes "over the function of the inactive gene at the (*pyr*) *3a* locus" as was suggested previously.[1] In fact it seems likely that the pyrimidine requirement of *pyr 3a* is not due to an "inactive gene," but rather to the activity of a gene being revised in a way such that, in a given genetic background, a pyrimidine requirement results. The requirement, the suppression of it and the other interactions as well, may result from alterations in the balance of reaction rates under the influence of the various mutant genes. The alterations may be pictured as arising through a variety of means, such as changes in enzyme concentrations or activities which produce changes in substrate or inhibitor levels affecting many reactions. The reactions involved may be related by a common coupled system or intracellular organization which is disturbed or modified in some way by the mutant genes. The variety of interactions observed does not point to any one mechanism as a very probable explanation, but rather, appears to suggest a greater number of possible mechanisms.

When the interactions involving the lysineless mutants were first reported it appeared to the authors that these, together with the observed accumulation of uracil by certain lysineless mutants,[2] indicated a close relationship between lysine and pyrimidine biosynthesis. In view of the further complexities reported here, a direct interrelation no longer seems necessarily indicated.

The shifting of the position of the apparent "genetic block" by modifiers has already been reported by Haskins and Mitchell[8] in a paper dealing with tryptophan-nicotinicless mutants. This phenomenon, which is represented here by the action of *s* and *pyr 3a* on the ornithineless, citrullineless and lysineless mutants, suggests that the relationship between the mutation and the nutritional requirement may not always be as direct as it has sometimes been considered to be. As Haskins and Mitchell[8] have suggested, the assumption that the gene affected by the mutation controls, in a direct fashion, the reaction which, on the basis of the results of growth experiments appears to be blocked, does not seem to be a safe one.

The fact that *pyr 3a* and its possible alleles, *3b* and *3c* are suppressed by *s*, whereas *pyr 1* and *2* are not, was at first considered by the authors to be evidence that *pyr 3a*, *b* and *c* were indeed alleles. At that time suppressors which acted on non-allelic mutants were known in Drosophila[15] and more recently Giles[16] has found that certain non-allelic methionineless mutants of Neurospora have a common suppressor and Lein and Lein[17] have reported that the same is true of three non-allelic acetateless mutants, also of Neurospora. It is, then, abundantly clear that response to the same suppressor cannot be considered as proof of allelism. Suppression by *s* of mutants having different requirements is reminiscent of the cases known in Drosophila[15] in which the same suppressor acts upon mutants which are not only non-allelic but which also have different and seemingly unrelated phenotypic effects.

* Work supported by grants from the Rockefeller Foundation.

[1] Houlahan, Mary B., and Mitchell, Herschel K., PROC. NATL. ACAD. SCI., **33,** 223 (1947).

[2] Houlahan, Mary B., and Mitchell, Herschel K., *Ibid.*, **34,** 465 (1948).

[3] Srb, A. M., and Horowitz, N. H., *J. Biol. Chem.*, **154,** 129 (1944).

[4] Srb, A. M. Botanical Gazette, **III**, 471 (1950).

[5] Srb, A. M., Fincham, J. R. S., and Bonner, D., *Am. J. Bot.*, **37,** 533 (1950).

[6] Haas, F., Mitchell, M. B., Ames, B. N., and Mitchell, H. K., *Genetics* (in press).

[7] Fincham, J. R. S., *J. Genetics*, **50,** 221 (1951).

[8] Haskins, F. A., and Mitchell, H. K. (in press).

[9] Beadle, G. W., and Tatum, E. L., *Am. J. Bot.*, **32,** 678 (1945).

[10] Mitchell, H. K., and Houlahan, M. B., *Fed. Proc.*, **6,** 506 (1947).

[11] Michelson, A. M., Drell, W., and Mitchell, H. K., PROC. NATL. ACAD. SCI., **37,** 396 (1951).

[12] Houlahan, M. B., Beadle, G. W., and Calhoun, H. G., *Genetics*, **34,** 493 (1949).

[13] Zalokar, M., PROC. NATL. ACAD. SCI., **34,** 32 (1948).

[14] Wagner, R. P., and Haddox, C. H., *Am. Nat.*, **85,** 319 (1951).

[15] Bridges, C. B., and Brehme, K., "The Mutants of *Drosophila melanogaster*," *Carnegie Inst. Washington Publ.* **552**, Washington, D. C., 257 pp. (1944).

[16] Giles, N. H., *Cold Spring Harbor Symposia.*

[17] Lein, P., and Lein, J., PROC. NATL. ACAD. SCI., **38,** 44 (1952).

THE EFFECTS OF GENE CHANGE ON TRYPTOPHAN DESMOLASE FORMATION*

BY CHARLES YANOFSKY

OSBORN BOTANICAL LABORATORY, YALE UNIVERSITY, NEW HAVEN, CONN.

Communicated by G. W. Beadle, December 29, 1951.

Genetic and biochemical investigations have provided substantial evidence for the view that physiological reactions are under the control of nuclear genetic material. These investigations have also indicated that the inability to carry out cellular biochemical reactions is a phenotypic expression of gene change and presumably results from gene-determined variations in the formation and/or the characteristics of particular enzyme systems.

Although the existing evidence suggests a close relationship between gene, enzyme and physiological reaction, the manner in which enzymes are controlled by genes and affected by gene change is poorly understood. In an effort to define more clearly the relationship between gene and enzyme, the present investigation was initiated.

Characteristics of Mutant Strains C-83 and S-1952.—The two Neurospora mutants employed in this investigation, strains C-83[1] and S-1952,[2] are characterized by a tryptophan requirement which cannot be satisfied by indole. The biochemical reaction presumably blocked in the two mutants, the coupling of indole and serine, has been demonstrated to be the principal and probably the sole means of tryptophan synthesis in Neurospora.[3] Since the enzyme catalyzing tryptophan synthesis, tryptophan desmolase, is extractable from wild type mycelium in an active cell-free state,[4] mutants C-83 and S-1952 constitute material in which the relationship of gene to enzyme can be investigated readily.

The tryptophan requirement characteristic of mutants C-83 and S-1952 was demonstrated to be due to a single gene. From crosses of each mutant with several standard wild type strains, invariably 1:1 segregation of the tryptophan requirement was observed in all asci tested. Both the C-83 and S-1952 mutant genes are approximately the same distance from the centromere (C-83 uncorrected centromere distance = 32 based on 51 asci, S-1952 uncorrected centromere distance = 33 based on 57 asci) and are linked to the fluffy locus (fluffy locus centromere distance = 29).[5] Allelism tests were performed with the two mutants by Dr. Dorothy Newmeyer of Stanford University and of 6300 ascospores obtained by crossing the mutants to each other, no tryptophan independent ascospores were recovered. Attempts in this laboratory to form tryptophan independent heterocaryons between the two mutants were also unsuccessful. Preliminary reversion experiments employing ultraviolet light as a mutagenic

agent indicate that both mutant genes revert to tryptophan independence (T^+) at about the same frequency (C-83, 43 $T^+/2.9 \times 10^7$ viable macroconidia; S-1952, 44 $T^+/1.8 \times 10^7$ viable macroconidia). Together these observations suggest that the C-83 and S-1952 mutant genes (henceforth designated as td_1 and td_2, respectively) are either allelic or closely linked. In the absence of contradictory evidence, they will be considered alleles in this paper. However, further experiments are in progress which are designed to examine more fully the linkage relationships between the two mutant genes and the nature of the reversions to tryptophan independence.

Since both mutants are blocked in the conversion of indole to tryptophan their culture filtrates were tested for possible indole accumulation. Indole (tentatively identified on the basis of color reactions with *p*-dimethylaminobenzaldehyde and xanthydrol) and anthranilic acid as well were present in culture filtrates from both strains. Mutant S-1952 was consistently found to accumulate several times as much indole as did mutant C-83. Whether or not this can be used as a distinguishing characteristic is difficult to tell at present since tryptophan dependent cultures obtained by crossing the two mutants to wild type strains show many different levels of indole accumulation. However, strains with the td_2 allele always accumulate more indole than do strains with the td_1 allele.

The nutritional behavior of mutants C-83 and S-1952 suggests that both strains lack the enzyme present in wild type strains which catalyzes tryptophan synthesis from indole and serine. To obtain direct evidence bearing on this possibility experiments were performed to determine whether or not extracts from mutant mycelium contain tryptophan desmolase activity. These tests were carried out with extracts and preparations from mutant mycelium which was grown under various culture conditions. The methods employed in obtaining tryptophan desmolase preparations are described in detail elsewhere.[6] For greater assay sensitivity a modification of the tryptophan desmolase assay procedure[6] was employed in these experiments. The incubation mixture contained 0.2 or 0.3 μM indole, 60 μM *DL*-serine, 20 μg. barium pyridoxal phosphate,[7] 0.06 ml. of 0.5 M phosphate buffer at pH 7.8 and enzyme extract in a 1 ml. volume. Assay tubes were generally incubated for two hours at 37°. At the end of this period 0.2 ml. of 5% NaOH was added to each tube to stop the reaction. Following the addition of base, 4 ml. of toluene were added to each tube. The tubes were then shaken and the layers allowed to separate. Aliquots removed from the toluene layer were assayed for indole,[8] and when desired, tryptophan was determined[9] in aliquots of the lower layer. The assay conditions employed are capable of detecting $^1/_{400}$ of the amount of tryptophan desmolase activity present in extracts from wild type strains.

Table 1 presents a summary of the tests carried out with mutant extracts and with various fractions obtained from these extracts. In addition to these tests, mutant extracts were incubated in the presence of pyridoxamine phosphate, yeast extract, casein hydrolysate, and with various

TABLE 1

PREPARATIONS FROM MUTANTS C-83 AND S-1952 EXAMINED FOR TRYPTOPHAN DESMOLASE ACTIVITY

STRAIN	GROWTH SUPPLEMENT	GROWTH TEMP.	AGE OF MYCELIUM, DAYS	METHOD OF EXTRACTION	FRACTIONS TESTED
C-83	Optimal T	25° 30°	2, 3, 4, 5 3, 4, 5	Homogenized in 0.1 *M* phosphate at pH 7.8	CE, DCE
C-83	Optimal T	30°	3	Homogenized in 0.1 *M* phosphate at pH 7.8	Cell debris CE, 30%, 50%, 70% $(NH_4)_2SO_4$ ppts.
C-83	Limiting T Optimal T Limiting T + CH Optimal T + CH	30°	3	Homogenized in 0.1 *M* phosphate at pH 7.8	CE, DCE, PP, DPP
C-83	Limiting T + S	25° shaken	4	Homogenized in 0.1 *M* phosphate at pH 7.8	CE, PP
C-83	Optimal T + S	30°	3	Mycelium extracted with water and assayed at various pH's	CE
C-83	Optimal T	30°	3	Extract of lyophilized mycelium	Cell debris CE, DCE
S-1952	Optimal T Limiting T Limiting T + S	30°	2, 3	Homogenized in 0.1 *M* phosphate at pH 7.8	CE, DCE, PP
S-1952	Limiting T + S	25° shaken	4	Homogenized in 0.1 *M* phosphate at pH 7.8	CE, PP
C-83-*su*	Optimal T Optimal T + S	30°	3	Homogenized in 0.1 *M* phosphate at pH 7.8	CE, DCE, PP, DPP

T = Trypotophan. S = Serine. CE = Crude extract. CH = Casein hydrolysate. DCE = Dialyzed crude extract. PP = The partially purified preparation obtained by fractionating a crude preparation according to the procedure used with wild type extracts. DPP = Dialyzed purified preparation.

concentrations of indole, serine and pyridoxal phosphate. Assays were also performed at several temperatures. Tryptophan desmolase activity could not be detected in mutant extracts under any conditions nor could activity be detected in any fraction obtained from these extracts. Mutant extracts were also mixed with wild type tryptophan desmolase prepara-

tions to test for the presence of inhibitors and no appreciable stimulation or inhibition was observed.

These results are in agreement with those obtained previously by Mitchell and Lein[10] in a study of one of these mutants, C-83. They too were unable to detect tryptophan desmolase activity in extracts from this mutant. The only report of the presence of tryptophan desmolase activity in mutant C-83[11] has been retracted[12] as a result of further experimentation. At the present time, therefore, it appears that neither extracts from strain C-83 nor from strain S-1952 contain tryptophan desmolase activity which can be detected by the methods employed.

Although tryptophan desmolase activity could not be demonstrated in mutant extracts, it is conceivable that there is active desmolase in the mutants and that the enzyme is destroyed during extraction or is inactive under the assay conditions employed. To obtain some data concerning

TABLE 2

CROSSES INVOLVING THE SUPPRESSOR GENE

	CROSS	ASCUS TYPE		
		$8T^+$	$6T^+:2T^-$	$4T^+:4T^-$
1.	Chilton wild type × S-1952 "reverted"	1	6	2
2.	Lein wild type × S-1952 "reverted"	1	8	3
3.	Chilton wild type carrying the suppressor × S-1952	1	8	2
4.	Lein wild type carrying the suppressor × S-1952	3	11	3
	Total	6	33	10
5.	Chilton wild type carrying the suppressor × C-83	0	0	5
6.	Lein wild type carrying the suppressor × C-83	0	0	5
	Total	0	0	10

T^+ = Tryptophan independent. T^- = Tryptophan dependent.

these possibilities an isotope experiment was performed to determine whether mutant C-83, when growing, is capable of tryptophan synthesis. The details of this experiment have been published previously.[13] In brief, mutant C-83 was grown on N^{15} ring labeled tryptophan and after maximum growth was attained tryptophan was isolated from the mycelium formed. The isolated tryptophan had the same N^{15} content as the labeled tryptophan fed, showing that mutant C-83 did not synthesize a significant amount of tryptophan under the conditions of this experiment.

These data, then, all support the view that mutants C-83 and S-1952, biochemically similar in their behavior, are both unable to synthesize tryptophan because of the absence of a normally essential enzyme, tryptophan desmolase. Thus these two mutants would appear to represent strains in which the result of gene mutation is the loss of a particular enzymatic activity.

Suppressor Studies.—During the course of this work a microconidial

stock of strain S-1952 reverted to tryptophan independence. Microconidia from this reverted culture were plated out and a homocaryotic tryptophan independent culture was recovered. This strain was then crossed with two standard wild type strains and as can be seen in table 2 (crosses 1 and 2), three types of asci were obtained. The most frequent ascus class had 6 tryptophan independent and 2 tryptophan dependent ascospores. This finding suggests that the reversion to tryptophan independence was due to a genetic change at a separate locus (*su*). This change appears to result in the suppression of the tryptophan requirement associated with the td_2 allele. As a further check of this conclusion tryptophan independents recovered from 4:4 asci ($4T^+:4T^-$) were crossed to mutant S-1952 and again three types of asci were obtained (table 2, crosses 3 and 4).

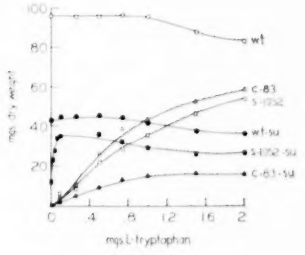

FIGURE 1

Growth of various strains of Neurospora on different levels of tryptophan. Incubated for 72 hours at 30°.

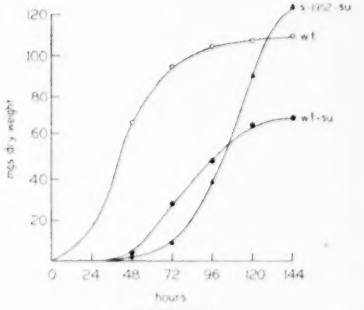

FIGURE 2

Growth curves of various strains on minimal medium at 30°.

The cultures obtained from these crosses were scored by inoculating two flasks with conidia, one flask containing minimal medium and the other flask containing minimal medium plus 1 mg. *L*-tryptophan. These flasks were incubated for 72 hours at 30° after which the mycelial pads were removed, dried and weighed. By employing this procedure it is possible to distinguish the four possible types of cultures: mutant, suppressed mutant, wild type and wild type carrying the suppressor gene. The mutant and wild type cultures can easily be identified since the former cultures do not grow on minimal medium while the latter cultures grow well on both types of media. The other two types of tryptophan independent cultures can be identified also, since the suppressed mutant grows poorly on minimal medium (in a 72-hour period) and better in the presence of tryptophan, while the growth of wild type strains carrying the suppressor gene is poor on the same two media and is not stimulated by

tryptophan. Comparisons of the growth of the various types of strains are shown in figures 1 and 2. In figure 2 it is seen that there is a lag in the growth of strains carrying the suppressor gene. Attempts to overcome this lag or inhibition associated with the suppressor gene by adding various supplements have been unsuccessful.

To obtain some information concerning the specificity of the suppressor gene, wild type strains carrying this gene were crossed to mutant C-83. As may be seen in table 2 (crosses 5 and 6) only 4:4 asci ($4T^+:4T^-$) were obtained. Thus the effect of the suppressor gene appears to be specific for the td_2 allele. That some of the tryptophan dependent cultures derived from these crosses actually carry the suppressor gene was demonstrated by performing crosses with strain S-1952. Additional crosses are being performed to determine if a particular genetic background is necessary for the suppressor gene to show its effect on the td_1 allele. Green[14] has also reported a case in which a suppressor gene does not affect every member of an allelic series

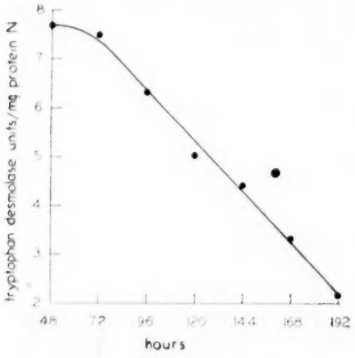

FIGURE 3

Variation in the tryptophan desmolase content of wild type mycelium with time. For the unit of enzyme activity see (6).

C-83 strains carrying the suppressor gene can be distinguished from C-83 strains without this gene since the presence of the suppressor gene results in a marked inhibition of growth (see Fig. 1). It can also be seen in figure 1 that C-83 strains require less tryptophan to reach maximum growth when they carry the suppressor gene. However, since the growth of strains with the td_1-*su* genotype is poor under all conditions, it is difficult to evaluate this observation.

Examination of Strains Carrying the Suppressor Gene for Tryptophan Desmolase Activity.—Since S-1952-*su* strains grow in the absence of exogenous tryptophan, enzyme studies were performed to see if this ability could be correlated with the presence of tryptophan desmolase. Also examined for tryptophan desmolase activity were two wild type strains carrying the suppressor gene and a C-83 strain carrying the suppressor gene (tryptophan dependent). Tryptophan desmolase activity was present in extracts from strain S-1952-*su* as well as in extracts from wild type strains carrying the suppressor gene. No activity could be detected in extracts from strain C-83-*su*. Attempts to activate extracts from this latter strain were also unsuccessful (see table 1). Thus the suppressor gene when present with

the td_2 allele restores both the ability to synthesize tryptophan desmolase and to grow in the absence of exogenous tryptophan; neither of these effects is observed in strains with the td_1-*su* genotype.

An effort was made to compare the amount of tryptophan desmolase in strain S-1952-*su* with the amount in wild type strains with and without the suppressor gene. The tryptophan desmolase content of wild type mycelium varies considerably with time as can be seen in figure 3 and, in view of the growth behavior of strains S-1952-*su* and wild type carrying the suppressor gene (see Fig. 2), such a comparison is difficult. Accepting these uncertainties, the maximum amount of tryptophan desmolase found in a suppressed S-1952 strain was approximately $^1/_{20}$ of the maximum amount of tryptophan desmolase that is formed by wild type strains.

TABLE 3

COMPARISON OF THE EFFECT OF VARIOUS INHIBITORS ON TRYPTOPHAN DESMOLASE FROM STRAIN S-1952-*su* AND FROM A WILD TYPE STRAIN

ADDITION	FINAL CONCENTRATION, MOLAR	% INHIBITION	
		SUPPRESSED MUTANT DESMOLASE	WILD TYPE DESMOLASE
Hydroxylamine	1×10^{-3}	100	92
F^-	1×10^{-3}	2	3
Co^{++}	1×10^{-3}	73	76
CN^-	1×10^{-3}	40	54
Zn^{++}	1×10^{-3}	78	81
Mg^{++}	1×10^{-3}	0	0
NH_4Cl	0.19	29	23
NH_4Cl	0.37	44	46
L-tryptophan	1×10^{-3}	22	19
L-tryptophan	2×10^{-3}	36	29

Thus the presence of the suppressor gene restores to strain S-1952 the ability to form effective concentrations of tryptophan desmolase but does not permit tryptophan desmolase synthesis to proceed to the same extent as it does in wild type strains. The stimulation of growth of strain S-1952-*su* by low concentrations of tryptophan (see Fig. 1) is in agreement with this finding. Wild type strains carrying the suppressor gene have a maximum tryptophan desmolase content which is about one-third that of wild type strains without the suppressor gene.

Comparison of Wild Type Tryptophan Desmolase and Tryptophan Desmolase from Strain S-1952-su.—Several characteristic properties of the tryptophan desmolase from a wild type strain were next compared with the properties of tryptophan desmolase from strain S-1952-*su*. On a molar basis indole uptake and tryptophan formation are equivalent in the presence of preparations from either strain. The stability of both preparations is roughly the same, both losing approximately 50% of their trypto-

phan desmolase activity when stored at 2° for 24 hours. Both preparations are inhibited by the same substances, and to practically the same extent (see table 3). As was the case with wild type preparations,[6] indole was not taken up by suppressed-mutant tryptophan desmolase preparations in the presence of the following substitutes for serine: alanine, glycine, cysteine, cystine, methionine, threonine, aspartate, glutamate, acetate, pyruvate, succinate or malate.

Before other comparisons were performed an attempt was made at fractionation and partial purification of suppressed-mutant tryptophan desmolase preparations according to the procedure developed for wild type preparations.[6] This procedure involves the use of various concentrations of ammonium sulfate at two pH values. Suppressed-mutant preparations and wild type preparations behaved identically throughout the steps in this procedure. The partially purified (about 20-fold) suppressed-mutant tryptophan desmolase preparations which were obtained were then employed in further comparisons. The pH-activity curves for such a suppressed-mutant preparation and for a wild type preparation are shown in figure 4. Both preparations exhibit maximum activity at pH 7.8.

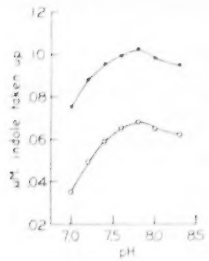

FIGURE 4

pH-activity curves for tryptophan desmolase preparations from a wild type strain (●—●) and from strain S-1952-*su* (○—○). 1 hour incubation at 37°. 2×10^{-4} *M* indole, 6×10^{-2} *M* DL-serine, 20 μg. barium pyridoxal phosphate per ml.

Comparisons of the affinity of the two partially purified enzyme preparations for their substrates and for their coenzyme were also performed. In the determinations of the Michaelis constant for serine, excesses of indole and pyridoxal phosphate were employed and indole uptake was measured. A Lineweaver-Burk plot[15] of kinetic data for serine is shown in figure 5. The approximate K_m value for serine for both preparations can be estimated from figure 5 to be 3.4×10^{-3} *M*. This value agrees fairly well with the value of 6×10^{-3} *M* found for wild type tryptophan desmolase preparations under different conditions.[6] Pyridoxal phosphate saturation curves are shown in figure 6. The pyridoxal phosphate concentration required to restore half-maximum activity is approximately the same for the two enzyme preparations, 3×10^{-6} *M*. It can also be seen from figure 6 that neither enzyme preparation (both were partially purified) was active in the absence of pyridoxal phosphate. The velocity of indole uptake by both preparations is essentially zero order at indole concentrations higher than 2×10^{-4} *M*.

The energy of activation was determined for partially purified wild type and suppressed-mutant preparations. Excesses of indole and

pyridoxal phosphate and a limiting concentration of serine were employed. Indole uptake was measured and the corresponding decrease in serine concentration was calculated and is plotted in figure 7 (*a*). The velocity constants obtained from these curves were then used to determine the energy of activation. The log k against $1/T$ plot is shown in figure 7 (*b*). In this figure it can be seen that the points obtained with the two enzyme preparations fall on the same line. The energy of activation, therefore, is approximately 12,000 cal./mole. in each case.

Thus the tryptophan desmolase preparations obtained from both strains appear to be similar by the criteria applied.

Discussion.—The data presented in this paper permit certain conclusions concerning the relationship between gene and enzyme. The studies with

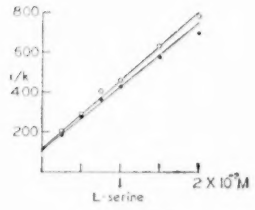

FIGURE 5

Lineweaver-Burk plot of kinetic data for serine. 1 hour incubation at 37°. Wild type tryptophan desmolase (●—●) S-1952-*su* tryptophan desmolase (○—○). 3×10^{-4} *M* indole, 20 μg barium pyridoxal phosphate per ml.

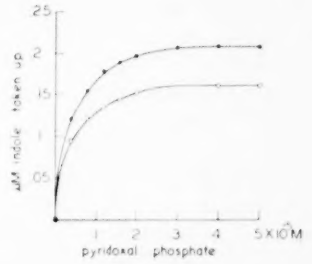

FIGURE 6

Pyridoxal phosphate saturation curves for tryptophan desmolase preparations from a wild type strain (●—●) and from strain S-1952-*su* (○—○). 1 hour incubation at 37°. 3×10^{-4} *M* indole, 6×10^{-2} *M* DL-serine.

mutants C-83 and S-1952 have shown that genetic alteration has affected the ability of each strain to produce active tryptophan desmolase. The nature of the effect on tryptophan desmolase or tryptophan desmolase formation which is responsible for the absence of detectable activity cannot be stated. It is clear, however, that if tryptophan desmolase is present in the mutants it does not provide sufficient tryptophan to support growth.

Although strains with the td_2 allele (S-1952) lack tryptophan desmolase activity, they are capable of forming active tryptophan desmolase when they are carrying a specific form of a different gene. This second gene, however, although it enables strains with the td_2 allele to form active enzyme, does not appear to be capable of mediating tryptophan desmolase

synthesis by itself. This conclusion is drawn from the observation that strains with the td_1-*su* genotype do not have tryptophan desmolase activity. In strain S-1952-*su* then, two genes, both in specific forms, appear to be required for the synthesis of tryptophan desmolase.

In addition to this case in which two specific genes appear to be required for the formation of a single enzyme, there are other cases in which mutation of any one of a number of genes affects the formation of single enzymes.[16-18] Together, these findings might be interpreted as indicating that the relationship between genes and enzymes is exceedingly complex. However, since so many cellular reactions are interrelated and interdependent it is difficult to tell at which level the complexity exists.

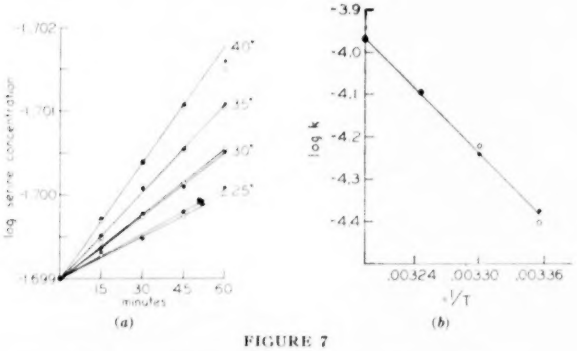

FIGURE 7

The effect of temperature on the rate of serine uptake (indole uptake was measured) by tryptophan desmolase preparations from a wild type strain (O—O) and from strain S-1952-su (●—●). 2×10^{-4} *M* indole, 4×10^{-2} *M* DL-serine, 30 μg. barium pyridoxal phosphate per ml.

In view of the fact that two genes, each differing from its wild type allele, are required for tryptophan desmolase synthesis in strain S-1952-*su*, one might expect the tryptophan desmolase formed by this strain to have altered properties. In a preliminary comparison in which a number of enzyme characteristics were examined, no major difference was found between the tryptophan desmolase extracted from strain S-1952-*su* and the same enzyme extracted from a wild type strain. To be sure, the methods employed could not reveal all types of changes that could occur in the suppressed mutant desmolase, nor are they sensitive enough to detect very slight changes. However, the catalytic portion of S-1952-*su* tryptophan desmolase, at least, does not appear to be altered appreciably.

The question of the effect of gene change on enzyme specificity has been

considered in other investigations[16–19] and, as yet, there is no clear demonstration of a change in enzyme specificity as a result of gene mutation. The recent findings of Pauling, *et al.*,[20] however, do suggest that such changes may occur. In any case, since relatively few studies of the effects of gene change on enzyme specificity have been performed, we cannot exclude alteration of enzyme specificity as a possible result of gene mutation.

It is difficult to account for the fact that the suppressor gene permits mutant S-1952 to grow in the absence of an exogenous supply of tryptophan and to form apparently normal tryptophan desmolase. The suppressor gene does not appear to be a duplication at another locus of the normal allele of the S-1952 locus since strains with the td_1-*su* genotype lack tryptophan desmolase activity. Nor does it seem likely that the suppressor gene acts by creating an alternate biosynthetic pathway of tryptophan formation or by relieving strain S-1952 from a tryptophan desmolase inhibitor. Although it is relatively easy to rule out most suggested explanations at the substrate level, at levels closer to the gene many hypotheses are consistent with the results obtained. Thus, the suppressor gene could act by creating an alternate biosynthetic pathway of tryptophan desmolase formation, by removing or inhibiting a reaction competing for a common enzyme precursor, by restoring the specificity of tryptophan desmolase, tryptophan desmolase precursors or the tryptophan desmolase forming system, by increasing the concentration of tryptophan desmolase precursors, tryptophan desmolase forming system or both or by affecting in any of a number of ways the rate of formation or time of appearance of tryptophan desmolase. It thus becomes apparent that we must have a better understanding of the processes involved in enzyme formation if we are to explain the mechanism of gene action.

The results obtained in this study also demonstrate that the td_1 and td_2 alleles behave differently. This can be seen from the fact that the suppressor gene, in contrast to its effect on strains with the td_2 allele, does not enable strains with the td_1 allele to grow in the absence of an exogenous supply of tryptophan, nor does it permit these strains to form detectable tryptophan desmolase. Thus the behavior of the td_1 and td_2 alleles in the presence of the suppressor gene suggests that these alleles may be qualitatively different forms of the same gene. In view of this and since there are several cases in which the complexity of single genetic loci has been demonstrated,[21–27] it is tempting to speculate that each gene may be composed of one or more units each of which is concerned with different steps in the synthesis of a single enzyme.

Summary.—Two allelic tryptophanless Neurospora mutants, both unable to use indole for growth, were examined for tryptophan desmolase, the enzyme which catalyzes tryptophan synthesis from indole in wild type strains. Tryptophan desmolase activity could not be detected in

extracts or in various preparations from either strain. One of these mutants, S-1952, does form tryptophan desmolase and grow in the absence of an exogenous supply of tryptophan when it is carrying a specific suppressor gene. This suppressor gene does not similarly affect the tryptophanless strain (C-83) with the allelic mutant gene. Since C-83-*su* strains do not form tryptophan desmolase, it appears that both the S-1952 mutant gene and the suppressor gene are required for tryptophan desmolase synthesis in strain S-1952-*su*. The tryptophan desmolase formed by a S-1952-*su* strain was compared with the tryptophan desmolase of a wild type strain and no major differences were detected.

Acknowledgment.—The author is greatly indebted to Dr. David M. Bonner for his guidance during the course of this investigation. The author is also indebted to Drs. G. W. Beadle, H. K. Mitchell and N. H. Horowitz of the California Institute of Technology for helpful suggestions concerning this manuscript.

* These investigations were supported in part by the Rockefeller Foundation and in part by the Atomic Energy Commission. [Contract No. AT (30-1)-1017].

[1] Strain C-83 was kindly supplied by Dr. H. K. Mitchell of the California Institute of Technology.

[2] Strain S-1952 was kindly supplied by Dr. E. L. Tatum of Stanford University.

[3] Tatum, E. L., and Bonner, D. M., PROC. NATL. ACAD. SCI., **30,** 30 (1944).

[4] Umbreit, W. W., Wood, W. A., and Gunsalus, I. C., *J. Biol. Chem.*, **165,** 731 (1946).

[5] Barratt, R. W., and Garnjobst, L., *Genetics*, **34,** 351 (1949).

[6] Yanofsky, C., *J. Biol. Chem.*, **194,** 279 (1952).

[7] Pyridoxal phosphate was kindly supplied by Dr. W. W. Umbreit of the Merck Institute for Therapeutic Research.

[8] Stanley, A. R., and Spray, R. S., *J. Bact.*, **41,** 251 (1941).

[9] Horn, M. J., and Jones, D. B., *J. Biol. Chem.*, **157,** 153 (1945).

[10] Mitchell, H. K., and Lein, J., *Ibid.*, **175,** 481 (1948).

[11] Gordon, M., and Mitchell, H. K., *Genetics*, **35,** 110 (1950).

[12] Horowitz, N. H., and Mitchell, H. K., *Ann. Rev., Biochem.*, **20,** 477 (1951).

[13] Partridge, C. W. H., Bonner, D. M., and Yanofsky, C., *J. Biol. Chem.*, **194,** 269 (1952).

[14] Green, M. M., *Records Genetic Society*, **20,** 101 (1951).

[15] Lineweaver, H., and Burk, D., *J. Am. Chem. Soc.*, **56,** 658 (1934).

[16] Markert, C., *Genetics*, **35,** 60 (1950).

[17] Lederberg, J., in *Genetics in the 20th Century*, Macmillan (1951).

[18] Landman, O., and Bonner, D. M., *Arch. of Biochem.*, in press.

[19] Lester, G., Ph.D. Dissertation, Yale University (1951).

[20] Pauling, L., Itano, H. A., Singer, S. J., and Wells, I. C., *Science*, **110,** 543 (1949).

[21] Green, M. M., and Green, K. C., PROC. NATL. ACAD. SCI., **35,** 586 (1949).

[22] Stadler, L. J., *Cold Spring Harbor Symposium XVI*, in press.

[23] Bonner, D. M., *Ibid.*, in press.

[24] Giles, N. H., *Ibid.*, in press.

[25] Roper, J. A., *Nature*, **166,** 956 (1950).

[26] Laughnan, J. R., *Records Genetics Society*, **19,** 108 (1950).

[27] Lewis, E. B., *Genetics*, **33,** 113 (1948).

PARTIAL DIFFERENTIAL EQUATIONS AND ANALYTIC CONTINUATIONS

BY S. BOCHNER

MATHEMATICS DEPARTMENT, PRINCETON UNIVERSITY

Communicated January 28, 1952

In Euclidean E^n we will say that an operator with constant coefficients

$$\Delta f = \sum_{p_1+\ldots+p_n\leq 2h} a_{p_1\ldots p_n} \frac{\partial^{p_1+\ldots+p_n} f}{\partial x_1^{p_1}\ldots\partial x_n^{p_n}} \tag{1}$$

of any even order is *elliptic semihomogeneous* (in a generalized sense) if it possesses a Green's function $G(x)$ which is defined and *analytic* in E^n except at the origin, and if there is an r_0 such that for any $r_1 \geq 0, \ldots, r_n \geq 0$, $r_1 + \ldots + r_n \geq r_0$ we have at infinity

$$\frac{\partial^{r_1+\ldots+r_n} G}{\partial x_1^{r_1}\ldots\partial x_n^{r_n}} = 0(|x|^{-n-1}), \qquad |x| \to \infty. \tag{2}$$

We have recently demonstrated[1] that such a Green's function exists if the polynomial

$$\sum_{p_1+\ldots+p_n\leq 2h} a_{p_1\ldots p_n}(it_1)^{p_1}\ldots(it_n)^{p_n}$$

and both its highest and lowest homogeneous parts occurring, are all three different from zero for any real $t_1, \ldots, t_n$ not all zero.

THEOREM 1. *In E_n, if D is a bounded domain having a connected boundary B, and if U is a neighborhood of B, and if an analytic function $f(x_j)$ in U satisfies an elliptic semihomogeneous equation*

$$\Delta f = 0 \tag{3}$$

in all its variables, and also an arbitrary (non-homogeneous, non-elliptic) equation

$$\Lambda f = \sum_{q_1+\ldots+q_m\leq M} b_{q_1\ldots q_m} \frac{\partial^{q_1+\ldots+q_m} f}{\partial x_1^{q_1}\ldots\partial x_m^{q_m}} = 0 \tag{4}$$

in fewer than all variables,

$$m < n,$$

the latter having again constant coefficients; then the function has an analytic continuation into all of $D + U$.

Proof: This theorem is an outgrowth of the theorem of Hartogs for complex variables, and it was already previously[2] given in the present version, and in more refined ones, but only for the operator (1) being the

ordinary Laplacean and the operator (4) being elliptic in its variables, more or less.

The Green's function of (1) gives rise to a Green's integral

$$g(x_j) = \int_B \sum_{\alpha=1}^{n} (X_\alpha \cdot \sigma^\alpha), \tag{5}$$

$$\sigma^\alpha = (-1)^{\alpha-1} d\xi_1 \ldots d\xi_{\alpha-1} d\xi_{\alpha+1} \ldots d\xi_n,$$

$$X_\alpha = \sum_{\mu_1+\ldots+\mu_n \leq 2h-1} \frac{\partial^{\mu_1+\ldots+\mu_n} f(\xi)}{\partial \xi_1^{\mu_1} \ldots d\xi_n^{\mu_n}} G^\alpha_{\mu_1 \ldots \mu_n}(x - \xi), \tag{6}$$

where (i) each $G^\alpha_{\mu_1 \ldots \mu_n}(\xi)$ is a linear combination with constant coefficients of derivatives of $G(\xi)$, (ii) for fixed x and any solution $f(\xi)$ of (3) we have

$$\frac{\partial X_1}{\partial \xi_1} + \ldots + \frac{\partial X_n}{\partial \xi_n} = 0 \tag{7}$$

and (iii) if B_0 bounds a domain in which $f(\xi)$ is a solution of (3) then for x in the domain we have $g(x_j) = f(x_j)$, that is

$$f(x_j) = \int_{B_0} \sum_\alpha X_\alpha \cdot \sigma^\alpha. \tag{8}$$

Now, for the proof of the theorem we may assume that B is a piecewise differentiable simplicial closed surface, and if we take a layer around it which is bounded by two "parallel" surfaces, and apply (8) to the layer then we obtain in U a Laurent decomposition

$$f(x_j) = f_1(x_j) - f_2(x_j), \tag{9}$$

where f_1 is analytic in $D + U$ and f_2 in $(E^n - D) + U$, and where for x in D, or in $E^n - D$, respectively, both are represented by the "same" integral (5). Our next step is to show that both are again solutions of the "accessory" equation (4),

$$\Lambda f_1 = 0, \qquad \Lambda f_2 = 0, \tag{10}$$

and as previously[2] we do this in the following manner. We view the integral (5) as an operator, obviously distributive, from solutions of the elliptic equation (3) in the neighborhood of B to functions $g(x)$ in $E^n - B$, denoting it then by $g(x) = L(f(\xi); x)$, and due to the fact that in (6) the variable x occurs only in the combinations $x - \xi$ this operator is commutative with partial differentiations as well, so that for any Λ with constant coefficients we have

$$\Lambda_x g(x) = L(\Lambda_\xi f(\xi); x). \tag{11}$$

If in particular $\Lambda f(\xi) = 0$ then also $\Lambda g(x) = 0$ and hence (10). It will be worth while to state the reason behind (11) in a comprehensive lemma, and we again introduce the differential form $\sigma^{\alpha\beta}$ which for $\alpha < \beta$ is defined as

$$(-1)^{\alpha+\beta-1}d\xi_1 \ldots d\xi_{\alpha-1}d\xi_{\alpha+1} \ldots d\xi_{\beta-1}d\xi_{\beta+1} \ldots d\xi_n,$$

for $\alpha = \beta$ is 0, and for $\alpha > \beta$ is $-\sigma^{\beta\alpha}$.

LEMMA 1. *If B is a (non-closed) surface with boundary C, and if in a neighborhood of $B + C$ we are given a solution $f(\xi)$ of an elliptic equation (3), and if we introduce the functional (11); then for any differential operator (4) with constant coefficients we have*

$$\Lambda_x g(x) - L(\Lambda_\xi f(\xi);\ x) = \int_C \sum_{\alpha,\beta=1}^{n} Y_{\alpha\beta}\sigma^{\alpha\beta},$$

where each $Y_{\alpha\beta}$ is a bilinear form in partial derivatives of $G(x - \xi)$, and in partial derivatives of $f(\xi)$ of order $\leq (2h - 1) + (M - 1)$.

For the conclusion of Theorem 1 it suffices to show that the function f_2 in (9) is analytic everywhere. We have already obtained $\Lambda f_2 = 0$, and from inspection of the integral (5) it follows that at infinity f_2 has the same behavior (2) as $G(x)$ itself, with another r_0 of course. Thus our conclusion follows from the following lemma which can be proved by Fourier analysis.

LEMMA 2. *If an analytic function $f(x_j)$ in an exterior of a sphere $|x| > R$ satisfies an equation (4) in fewer than all variables, and at infinity has the behavior (2), then $f(x_j)$ is a polynomial of degree $\leq M - 1$ in all its variables and is thus analytic everywhere.*

For the proof of the Laurent decomposition (9) instead of applying (8) to a layer around all of B we might have used as previously[2] the following "local" argument.

LEMMA 3. *If we consider the integral (5) for any union of simplices B and any $f(\xi)$ on it, and if in a neighborhood of point of a simplex, $f(\xi)$ is a solution of (3), then in the neighborhood of this point the function $g(x)$ can be continued analytically across B from both sides, and the difference of $g(x)$ on both sides is $f(x)$ itself,*

$$g_+(x) - g_-(x) = f(x).$$

With this lemma the following sharpening of Theorem 1 ensues.

THEOREM 2. *If the piecewise differentiable surface B is a union of the closures of differentiable simplices $B_1, \ldots, B_r$; if in a neighborhood U_ρ of each (non-closed) B_ρ there is a given analytic solution f_ρ of (3) and (4); if each f_ρ together with its partial derivatives of order $\leq (2h - 1) + (M - 1)$ has continuous boundary values on the boundary of B_ρ; and if on each $(n - 2)$-dimensional edge common to two simplices all these boundary values are the same for the two functions; then in the domain D bounded by B there exists an analytic function $f(x)$ which in each U_ρ coincides with the function $f_\rho(x)$ given.*

Another conclusion is as follows.

THEOREM 3. *If a surface B with a boundary C does not decompose E^n, and if in a neighborhood of B there is given a solution of (3) and (4) for which*

the partial derivatives (or only their boundary values) of order $\leq (2h - 1) + (M - 1)$ *are 0 on C; then* $f(x)$ *itself is* $= 0$.

A full account with results in other directions will be given elsewhere.

[1] "Zeta Functions and Green's Functions for Linear Partial Differential Operators of Elliptic Type with Constant Coefficients," *Ann. Math.* (to appear).

[2] "Analytic and Meromorphic Continuation by Means of Green's Formula," *Ibid.*, **44**, 652-673 (1943).

THE DIRICHLET PROBLEM FOR LINEAR ELLIPTIC EQUATIONS OF ARBITRARY EVEN ORDER WITH VARIABLE COEFFICIENTS

By Felix E. Browder

Boston University

Communicated by Marston Morse, January 22, 1952

Let K be an arbitrary linear elliptic partial differential operator of order $2m$ on a bounded domain D of Euclidean n-space. The Dirichlet problem for K on D asks for a solution u of the equation $Ku = 0$ in D such that u and its derivatives of order less than m assume the same values on the boundary of D, in a suitable sense, as a prescribed function g and its corresponding derivatives. The coefficients of K being suitably differentiable, it is shown that if the Dirichlet problem for K has at most one solution, then it has exactly one for every boundary value problem. More generally the boundary value functions for which a solution to the Dirichlet problem exists are characterized in terms of the null solutions of the adjoint equation by consideration of the inhomogeneous equation. Finally a theory of eigenfunctions and eigenvalues is established for the eigenfunctions of a large class of self-adjoint equations.

The proofs combine essential features of the method of orthogonal projection with the theory of completely continuous linear operators on a Hilbert space. They involve no assumptions of self-adjointness or positivity on K in contrast with the usual proofs based on the method of orthogonal projection or the Dirichlet principle.

These results are a generalization of previous general existence theorems based on variational methods including those of Courant, Friedrichs, and Rellich for certain classes of self-adjoint second-order linear elliptic equations,[1] Friedrichs for the biharmonic equation,[2] Soboleff and Vischik for the polyharmonic equations,[3] Kodaira for the positive homogeneous self-adjoint second-order linear elliptic equation,[4] and Gårding for general linear elliptic equations of arbitrary order with constant coefficients.[5]

1. If j is a positive integer, $C^j(D)$ is the set of j-times continuously

differentiable functions on D, $C_c^j(D)$ the subset of $C^j(D)$ of functions vanishing outside a compact subset of D.

For $f \in C_c^j(D)$, $g \in C^j(D)$ we define a j-inner product by

$$(f, g)_j = \int_D \left(\sum_{i_1 \ldots i_j} \frac{\partial^j f}{\partial x_{i_1} \ldots \partial x_{i_j}} \frac{\partial^j g}{\partial x_{i_1} \ldots \partial x_{i_j}} \right) dx$$

and a j-norm $\|f\|_j = (f, f)_j^{1/2}$. More generally the j-norm may be defined for any $g \in C^j(D)$ whose jth derivatives are in $L^2(D)$. The o-norm and inner product are those of $L^2(D)$.

There exist constants $c_j > 0$ for each $j < m$ for which $\|f\|_j \leq c_j \|f\|_{j+1}$ for all $f \in C_c^m(D)$.

If H is the completion of $C_c^m(D)$ with respect to the m-inner product, the Hilbert space H may be identified with a linear subset of $L^2(D)$. Thus identified, H will be designated as the set of functions of $L^2(D)$ which vanish with all their derivatives of order less than m on the boundary of D.

The Dirichlet Problem: Given $g \in C^m(D)$ for which $\|g\|_m < \infty$, to find $u \in C^{2m}(D)$ such that $Ku = 0$ in D and u-g and its derivatives of order less than m vanish on the boundary of D.

K is said to be suitably differentiable if the coefficients in K of the jth order derivatives belong to $C^{2m+j}(D)$ and if the mth derivatives of all the coefficients are continuous on the closure of D.[6] The formal adjoint $\bar{K}$ of K is well defined for such K.

K may be written in the form

$$K = \sum_{\substack{i_1 \ldots i_m \\ k_1 \ldots k_m}} p_{i_1 \ldots i_m k_1 \ldots k_m}(x) \frac{\partial^{2m}}{\partial x_{i_1} \ldots \partial x_{i_m} \partial x_{k_1} \ldots \partial x_{k_m}} + R$$

where the indices $i_1, \ldots, i_m, k_1, \ldots, k_m$ range independently from 1 to n, the functions $p_{i_1 \ldots i_m k_1 \ldots k_m}(x)$ are symmetric in the indices, and R is a differential operator of order less than $2m$.

K will be termed elliptic if both of the following two conditions are fulfilled:

(*a*) The form

$$\sum_{\substack{i_1 \ldots i_m \\ k_1 \ldots k_m}} p_{i_1 \ldots i_m k_1 \ldots k_m}(x) \xi_{i_1} \ldots \xi_{i_m} \xi_{k_1} \ldots \xi_{k_m}$$

is positive definite in the ξ-variables for each x in D.

(*b*) If we define (f, h) for $f, h \in C_c^m(D)$ by

$$(f, h) = \int_D \left(\sum p_{i_1 \ldots i_m k_1 \ldots k_m}(x) \frac{\partial^m f}{\partial x_{i_1} \ldots \partial x_{i_m}} \frac{\partial^m h}{\partial x_{k_1} \ldots \partial x_{k_m}} dx \right)$$

then there exists a constant $c > 0$ such that $(f, f) \geq c(f, f)_m$ for all $f \in C_c^m(D)$

If K is of the second-order or if the terms of highest order have constant coefficients then condition (*b*) follows from (*a*) which is the usual ellipticity condition.

THEOREM 1. *If K is a suitable differentiable linear elliptic differential operator on a bounded domain D in E^n and if the Dirichlet problem for zero boundary values has only the zero solution, then the Dirichlet problem has a solution for every boundary function g.*

An important part in the proof of Theorem 1 is played by the following generalization of the well-known Weyl Lemma.

THEOREM 2. *Let K be a linear elliptic operator for which the coefficients of the jth derivatives belong to $C^{2m+j}(D)$. Suppose that v is a measurable function on D square summable on every compact subset of D, $\psi \in C^1(D)$. Suppose also that for all $f \in C_c^{2m}(D)$*

$$\int_D v\bar{K}f\,dx = \int_D \psi f\,dx.$$

Then v is equal almost everywhere in D to u where $u \in C^{2m}(D)$ and $Ku = \psi$ in D.

The proof of Theorem 2 is based on a method which has been applied under various hypotheses by H. Weyl,[7] Kodaira,[4] De Rham and Kodaira,[8] L. Schwartz[9] and L. Gårding.[10] The existence of a fundamental solution in the small for linear elliptic equations with analytic coefficients as established by F. John[11] is used in the proof.

Proof of Theorem 1: By the Schwarz inequality there exists a constant $d > 0$ for which

$$\|f\| \leq d\|f\|_m$$

for all $f \in C_c^m(D)$ where $\|f\| = (f, f)^{1/2}$. Thus (f, h) may be extended to an inner product on H under which H is a Hilbert space with a norm equivalent to the m-norm.

Let L be the self-adjoint linear elliptic differential operator defined by

$$L = \sum_{\substack{i_1 \ldots i_m \\ k_1 \ldots k_m}} \frac{\partial^m}{\partial x_{i_1} \ldots \partial x_{i_m}} \left(p_{i_1 \ldots i_m k_1 \ldots k_m}(x) \frac{\partial^m}{\partial x_{k_1} \ldots \partial x_{k_m}} \right)$$

Then $K = L + A$, A of order less than $2m$.

Consider (h, g) for h in H and the given boundary value function g. By the Schwarz inequality

$$|(h, g)| \leq a\,\|h\|_m \|g\|_m,\ a > 0$$

and hence (h, g) is a continuous linear functional for h in H. Therefore there exists $h_0 \in H$ such that $(h, h_0) = (h, g)$ for all h in H.

Set $v = g - h_0$. Then $v - g = -h_0 \in H$ and $(h, v) = 0$ for all h in H. In particular if $f \in C^{2m}(D)$,

$$0 = (f, v) = (-1)^m \int_D vLf\,dx$$

By Theorem 2 there exists $v_0 \in C^{2m}(D)$ for which $Lv_0 = 0$ and $v_0 - g \in H$. In addition $\|v_0\|_m < \infty$.

The existence of a solution u of the equation $Ku = 0$ assuming the boundary values of g is equivalent with the existence of a solution w of the equation

$$Kw = K(u - v_0) = -Av_0$$

with zero boundary values. By Theorem 2, w belongs to $C^{2m}(D) \cap H$ and is a solution of the above equation if and only if

(1) $\int_D w\bar{K}f\,dx = -\int_D Av_0 f\,dx$ for all $f \in C_c^{2m}(D)$.

(1) can be written in the form

(2) $\int wLf\,dx + \int w\,A\,f\,dx = \int Av_0 f\,dx$.

For $w \in H$, $\int_D wLf\,dx = (-1)^m(w, f)$.

Secondly, there exists $\rho_1 > 0$ for which

$$\left| -\int_D Av_0 f\,dx \right| \leq \rho_1 \|f\|_m$$

for all $f \in C_c^{2m}(D)$. As a consequence there exists $\eta \in H$ such that

$$-\int Av_0 f\,dx = (\eta, f)$$

for all $f \in C_c^{2m}(D)$. Finally it may be shown by a simple argument that there exists a constant $\rho_2 > 0$ for which

$$\left| \int w\bar{A}f\,dx \right| \leq \rho_2 \|w\|_{m-1} \|f\|_m \leq \rho_2 c_{m-1} \|w\|_m \|f\|_m$$

It follows from these last inequalities and the Rellich Selection Theorem[12] that a subset of H bounded in the m-norm is conditionally compact in the m-1-norm and consequently that there exists a completely continuous linear mapping V of H into itself for which

$$(Vw, f) = \int_D w\bar{A}f\,dx$$

for all $f \in C_c^{2m}(D)$. Thus (2) can be written as

$$((-1)^m w + Vw - \eta, f) = 0$$

for all $f \in C_c^{2m}(D)$.

A necessary and sufficient condition that this latter equation hold is that

$$w + (-1)^m Vw = (-1)^m \eta$$

for some $w \in H$. $(-1)^m V$ is a completely continuous linear mapping of H into itself and by a theorem of F. Riesz[13] on completely continuous linear transformations of a Banach space $I + (-1)^m V$ is onto if the equation

$$w + (-1)^m Vw = 0$$

has no solution in H except $w = 0$.

Any such null solution, however, would also be a solution of the equation

$$0 = (-1)^m(w, f) + (Vw, f) = \int_D wLf\,dx + \int_D w\bar{A}f\,dx = \int_D w\bar{K}f\,dx.$$

Thus

$$\int_D w\bar{K}f\,dx = 0$$

for all $f \in C_c^{2m}(D)$ and, by Theorem 2, w which belongs to H would be almost everywhere equal to a $2m$-times differentiable solution w_0 of the equation $Kw_0 = 0$. But w_0 is a solution of the Dirichlet problem with null boundary values and must be identically zero. Hence $w = 0$ almost everywhere in D and is the 0 element of H. Therefore the Dirichlet problem for K on D with boundary value function g has a solution. Q.E.D.

COROLLARY TO THEOREM 1: *Let K be a suitably differentiable linear elliptic differential operator on a bounded domain D in E^n. Suppose that there exists a constant $\rho_3 > 0$ such that*

$$(-1)^m \int Kf \cdot f\,dx \geq \rho_3 \|f\|_m$$

for all $f \in C_c^{2m}(M)$. Then the Dirichlet problem for K has one and only one solution for every boundary value function g. In particular the above condition holds for a homogeneous self-adjoint differential operator like the L of the proof of Theorem 1.

THEOREM 3. *Let K be a suitably differentiable linear elliptic operator on a bounded domain D in E^n, $\psi \in L^2(D) \cap C^1(D)$. The equation $Ku = \psi$ has a solution $u \in C^{2m}(D)$ with zero boundary values if and only if*

$$\int_D \psi \cdot \varphi\,dx = 0$$

for each null solution φ of $\bar{K}\varphi = 0$. The linear dimension of the set of null solutions for $\bar{K}$ is finite and equal to that for K.

The methods developed in this note can be used to treat the eigenfunctions and eigenvalues of the Dirichlet problem for K. With suitable restrictions on the coefficients at infinity the theorems may be extended to unbounded domains.

[1] Courant-Hilbert, v. 2, Chap. VII.

[2] *Math. Ann.*, **98** (1927).

[3] *Math. Sbornik*, **2**, 44 (1937); *Ibid.*, **25** (1949).

[4] *Proc. Imp. Acad. Tokyo*, **20** (1944).

[5] *Compt. rend. Paris*, **230** (1950).

[6] The second condition though not the most economical possible is introduced for the sake of simplicity.

[7] *Duke Math. J.*, **7** (1940).

[8] Harmonic Integrals, Notes on Lectures at the Institute for Advanced Study, 1950, pp. 36–39.

[9] *Theorie des Distributions*, v. I, Paris, 1950, pp. 136 ff.

[10] *Proc. Roy. Physiog. Soc. Lund.*, **20**, 250–253 (1950).
[11] *Comm. on Pure and Applied Math.*, v. 3 (1950).
[12] Courant-Hilbert, v. 2, pp. 488–490.
[13] *Acta Math.*, **41** (1918).

---

*INTEGRATION OF STIFF EQUATIONS**

BY C. F. CURTISS AND J. O. HIRSCHFELDER

THE NAVAL RESEARCH LABORATORY, DEPARTMENT OF CHEMISTRY, UNIVERSITY OF WISCONSIN, MADISON, WISCONSIN

Communicated by Farrington Daniels, December 29, 1951

In the study of chemical kinetics, electrical circuit theory, and problems of missile guidance a type of differential equation arises which is exceedingly difficult to solve by ordinary numerical procedures. A very satisfactory method of solution of these equations is obtained by making use of a forward interpolation process. This scheme has the unusual property of singling out and approximating a particular solution of the differential equation to the exclusion of the manifold of other solutions. This behavior may be explained by a simple geometrical interpretation of the significance of the forward interpolation process. The differential equations to which this method applies are called "stiff."

A typical example of a stiff equation is the equation representing the rate of formation of free radicals in a complex chemical reaction. The free radicals are created and destroyed so rapidly compared to the time scale for the over-all reaction that to a first approximation the rate of production is equal to the rate of depletion. This is the notion of the pseudo-stationary state. In some cases such as the fast reactions occurring in flames or detonations, this approximation is not sufficiently accurate. The method described in the present paper provides a means for obtaining solutions to equations of this type to any degree of accuracy.

The numerical procedure described here can easily be extended to sets of simultaneous first-order differential equations. In any particular region, the differential equations can be uncoupled by introducing suitable linear combinations of the original dependent variables. Some of the uncoupled equations may be "stiff" in which case they can be integrated by the methods discussed here; whereas other uncoupled equations may be integrated by the more usual procedures.

1. Concept of Stiff Equations.—Consider the first-order differential equation,

$$\frac{dy}{dx} = [y - G(x)]/a(x, y). \tag{1}$$

The right-hand side of this equation represents a general function of x and y which for each value of x has a root, $y = G(x)$. If Δx is the desired resolution of x or the interval which will be used in the numerical integration, the equation is "stiff" if

$$\left|\frac{a(x, y)}{\Delta x}\right| \ll 1 \tag{2}$$

and $G(x)$ is well behaved [i.e., varies with x considerably more slowly

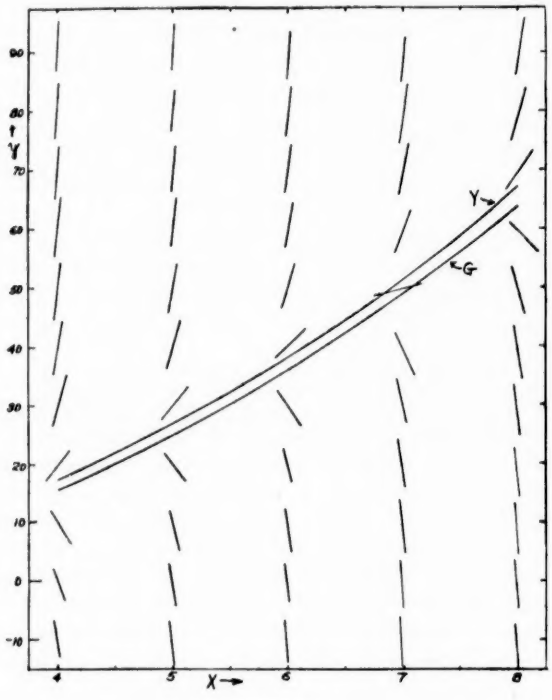

FIGURE 1

Slopes, dy/dx, for a Typical Stiff Equation, $dy/dx = 5(y - x^2)$.

than does $\exp(x/a(x, G(x)))]$. The geometrical significance of the stiffness is shown in figure 1. Here the family of solutions horn out as one proceeds in the positive x direction. This can be chosen as the positive x direction without loss of generality. The slope, dy/dx, is drawn at regular intervals of x and y. For a particular value of x: for large values of y the slope is very large and positive; for small values of y the slope is very large in the

negative sense; in the vicinity of $y = G(x)$ the slope changes rapidly from large positive to large negative.

Looking at equation (1), it appears that if $a(x, y)$ is sufficiently small there is a special solution, $y = Y(x)$, which lies close to $y = G(x)$. To a first approximation, Y is given by $Y^{(1)}$,

$$Y^{(1)} = G + a(x, G)\frac{dG}{dx}. \tag{3}$$

The second approximation, $Y^{(2)}$, is obtained by substituting $Y^{(1)}$ for y in dy/dx and in $a(x, y)$,

$$Y^{(2)} = G + a(x, Y^{(1)})\frac{dY^{(1)}}{dx}. \tag{4}$$

In this manner, an arbitrary order of approximation of Y may be obtained in terms of the lower orders of approximation. If $a(x, y)$ does not depend upon y, we can write down the resulting expression for Y,

$$Y = \sum_{k=0}^{\infty} \frac{D^k G}{Dx^k} \tag{5}$$

where D/Dx is the differential operator, $a(x)d/dx$. Whereas the function $Y(x)$ remains close to $G(x)$, every other solution deviates exponentially. This can be seen by subtracting the differential equation for Y from equation (1). In the region about $Y(x)$ (assuming that $a(x, y)$ varies slowly with y),

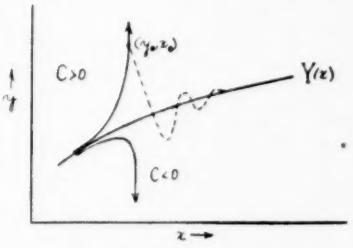

FIGURE 2

$$a(x, Y(x))\frac{d(y - Y)}{dx} = (y - Y). \tag{6}$$

This integrates to give

$$y - Y = C \exp\left[\int \frac{dx}{a(x, Y(x))}\right] \tag{7}$$

where C is a constant of integration. The qualitative appearance of this family of solutions is shown in figure 2.

2. *Numerical Solution of Stiff Equations Using Forward Interpolation.*— It is very difficult to integrate "stiff" equations by ordinary numerical methods. Small errors are rapidly magnified if the equations are inte-

grated in the direction such that the family of solutions horn out, whereas the numerical solutions oscillate violently about $Y(x)$ if the integration is carried out in the opposite direction.

We have discovered a method of numerically integrating "stiff" equations which has the desirable property that regardless of the starting conditions and regardless of the direction of integration, a solution is generated which rapidly converges to $Y(x)$. Thus for example, in figure 2, if we use this system of integration starting at the point (x_0, y_0) we obtain

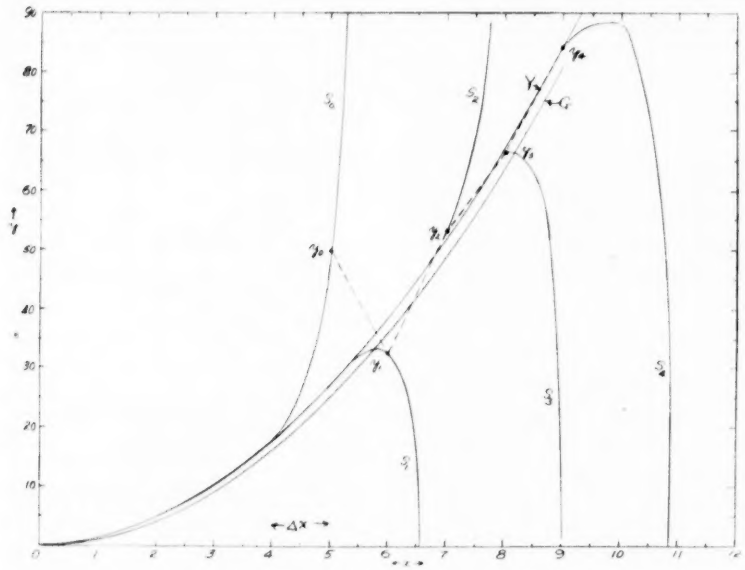

FIGURE 3

Integration of $dy/dx = 5(y - x^2)$ in the positive x direction from $(x = 5, y = 50)$. Here $a/\Delta x = 1/5$, $G(x) = x^2$, $Y(x) = 0.08 + 0.4x + x^2$.

the function indicated by the dotted line rather than the solution to the original differential equation which passes through this point and rapidly approaches infinity.

Let $x_0, x_1, \ldots, x_n, \ldots$ be a set of values of x spaced a distance Δx between successive points. Then the subscript on any quantity indicates that it is evaluated at the corresponding value of x. We wish to evaluate y_n from a knowledge of y at the previous points. Suppose, for example, that $y(x)$ can be approximated locally by a straight line passing through y_n and y_{n-1}. Then this straight line has the slope:

$$\left(\frac{dy}{dx}\right)_n = \frac{y_n - y_{n-1}}{\Delta x}. \tag{8}$$

Now, using this expression for the slope in evaluating the terms of the differential equation, equation (1), at the forward point one obtains:

$$\left(\frac{dy}{dx}\right)_n = \frac{y_n - y_{n-1}}{\Delta x} = \frac{y_n - G_n}{a(y_n, x_n)}. \tag{9}$$

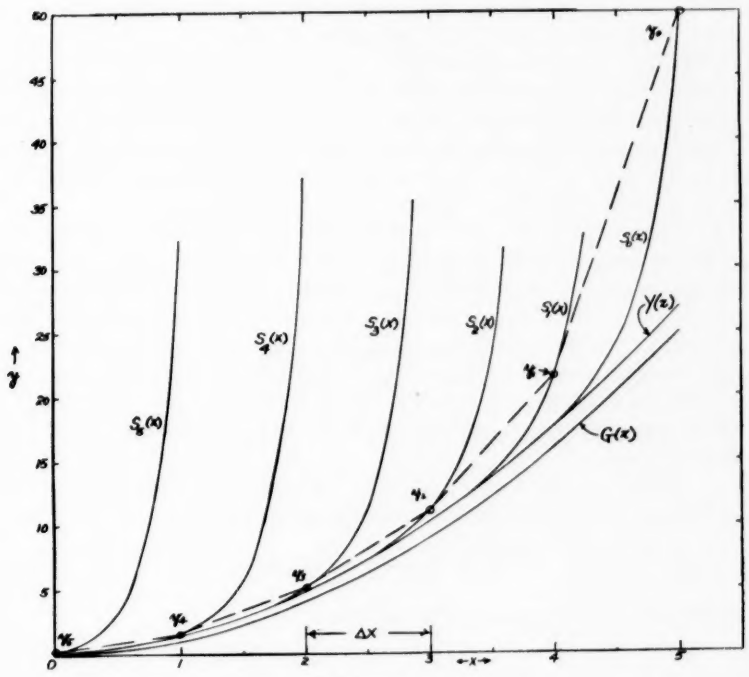

FIGURE 4

Integration of $dy/dx = 5(y - x^2)$ in the negative direction from $(x = 5, y = 50)$. Here $a/\Delta x = -1/5$, $G(x) = x^2$, $Y(x) = 0.08 + 0.4x + x^2$.

Thus,

$$y_n = \frac{G_n - \dfrac{a(x_n, y_n)}{\Delta x}\, y_{n-1}}{1 - \dfrac{a(x_n, y_n)}{\Delta x}}. \tag{10}$$

If $a(x, y)$ does not depend upon the value of y, equation (10) gives an explicit solution for y_n in terms of y_{n-1}; otherwise, equation (10) gives an implicit relationship between y_n and y_{n-1}. Thus starting from a point (x_0, y_0), equation (10) provides a numerical solution to the differential equation, equation (1). The curious fact is that quite regardless of the starting point, thus numerical solution seeks and approximates $Y(x)$ in the manner shown by the dotted curve in figure 2. The reason for this behavior may be seen geometrically from figure 3. Here y_0 lies on the curve $y = S_0(x)$ where $S_0(x)$ is a solution to the original differential equation. Equating $(dy/dx)_1$ from equations (8) and (9) is equivalent to asking what solution to equation (1), $y = S_1(x)$, has a tangent at the point x_1 which passes through the point (x_0, y_0). Similarly in passing from x_1 to x_2, etc. Since the slope for a stiff equation has a reasonable value (neither tremendously large in the positive or negative sense) only in the vicinity of $G(x)$ or $Y(x)$, it is clear that this numerical integration scheme limits us to approximating $Y(x)$.

It is interesting to notice that this method of integration may be used in either direction. An integration in the negative direction so that Δx is negative (i.e., in the direction of convergence of the manifold of solutions) is illustrated in figure 4. In this case the solution approaches the asymptote without oscillations. The integration in the two directions is not reversible since the definition of the forward point depends upon the direction of integration. This fact may be used numerically to hunt the special solution by first integrating in one direction and then reversing the direction.

3. The Asymptotic Form.—If $a(x, y)$ is a function only of x, it is easy to see how the numerical procedure leads to an approximation of Y. Since

$$G_n = Y_n - a(x_n)(dY/dx)_n, \tag{11}$$

equation (10) can be rewritten in the form

$$y_n - Y_n = -\left(\frac{\dfrac{a(x_n)}{\Delta x}}{1 - \dfrac{a(x_n)}{\Delta x}}\right)[y_{n-1} - Y_{n-1} + \epsilon] \tag{12}$$

where

$$\epsilon = -Y_n + Y_{n-1} + \Delta x \left(\frac{dY}{dx}\right)_n. \tag{13}$$

Expanding Y and (dY/dx) in Taylor series about the point, x_{n-1}, it follows that

$$\epsilon = \frac{1}{2}(\Delta x)^2\left(\frac{d^2Y}{dx^2}\right)_{n-1} + \frac{1}{3}(\Delta x)^3\left(\frac{d^3Y}{dx^3}\right)_{x_{n-1}} + \ldots. \qquad (14)$$

Looking at equation (12) it appears that there are two reasons why y_n differs from Y_n. The rate of convergence of y toward Y is measured by the factor

$$b = \frac{\dfrac{a(x_n)}{\Delta x}}{1 - \dfrac{a(x_n)}{\Delta x}}. \qquad (15)$$

The larger the interval size Δx, the smaller the value of b and the more rapidly y_n converges toward Y_n. The series converges for values of b less than unity and approaches an asymptote given approximately by

$$y = Y + a(x)\left[\frac{1}{2}(\Delta x)\frac{d^2Y}{dx^2} + \frac{1}{3}(\Delta x)^2\frac{d^2Y}{dx^3} + \ldots\right]. \qquad (16)$$

This discrepancy between y and Y is due to the poorness of the fit of Y by successive linear segments. The smaller the intervals, Δx, the smaller this error.

The error in the asymptote of the series may be reduced by using a three-point formula for $(dy/dx)_n$ in which y is fit to a quadratic passing through y_n, y_{n-1} and y_{n-2}. Since the slope is evaluated at x_n, the same sort of forward interpolation is used as in the linear approximation. In place of equation (10), one obtains the following relation:

$$y_n = \frac{G_n - 2\dfrac{a(x_n, y_n)}{\Delta x}y_{n-1} + \dfrac{1}{2}\dfrac{a(x_n, y_n)}{\Delta x}y_{n-2}}{1 - \dfrac{3}{2}\dfrac{a(x_n, y_n)}{\Delta x}}. \qquad (17)$$

In this case the difference between y_n and Y_n is given by[1]

$$y_n - Y_n = \left(\frac{\dfrac{a(x_n)}{\Delta x}}{1 - \dfrac{3}{2}\dfrac{a(x_n)}{\Delta x}}\right) \times \left[\begin{array}{l} -2(y_{n-1} - Y_{n-1}) + \dfrac{1}{2}(y_{n-2} - Y_{n-2}) \\ -\dfrac{1}{3}(\Delta x)^3\left(\dfrac{d^3Y}{dx^3}\right)_{n-1} - \dfrac{1}{12}(\Delta x)^4\left(\dfrac{d^4Y}{dx^4}\right)_{n-2} + \ldots \end{array}\right] \qquad (18)$$

Hence the asymptotic form is approximately

$$y = Y + a(x)\left[\frac{2}{9}(\Delta x)^2\frac{d^3Y}{dx^3} + \frac{1}{18}(\Delta x)^3\frac{d^4Y}{dx^4} + \ldots\right]. \quad (19)$$

A geometrical argument of why y_n approximates Y_n follows the same line as for the linear case.

The use of a cubic to fit y leads to the expression

$$y_n = \frac{G_n + \dfrac{a(x_n, y_n)}{\Delta x}\left(-3y_{n-1} + \dfrac{3}{2}y_{n-2} - \dfrac{1}{3}y_{n-3}\right)}{1 - \dfrac{11}{6}\dfrac{a(x_n, y_n)}{\Delta x}} \quad (20)$$

with a discrepancy between y_n and Y_n given by[2]

$$y_n - Y_n = \left(\frac{\dfrac{a(x_n)}{\Delta x}}{1 - \dfrac{11}{6}\dfrac{a(x_n)}{\Delta x}}\right) \times \left[\begin{array}{l} -3(y_{n-1} - Y_{n-1}) + \dfrac{3}{2}(y_{n-2} - Y_{n-2}) - \dfrac{1}{3}(y_{n-3} - Y_{n-3}) \\ -\dfrac{1}{4}(\Delta x)^4\left(\dfrac{d^4Y}{dx^4}\right)_{n-1} + \dfrac{1}{20}(\Delta x)^5\left(\dfrac{d^5Y}{dx^5}\right)_{n-1} + \ldots \end{array}\right]. \quad (21)$$

The use of a quartic leads to

$$y_n = \frac{G_n + \dfrac{a(x_n, y_n)}{\Delta x}\left(-4y_{n-1} + 3y_{n-2} - \dfrac{4}{3}y_{n-3} + \dfrac{1}{4}y_{n-4}\right)}{1 - \dfrac{25}{12}\dfrac{a(x_n, y_n)}{\Delta x}} \quad (22)$$

with the discrepancy between y_n and Y_n given by

$$y_n - Y_n = \left(\frac{\dfrac{a(x_n)}{\Delta x}}{1 - \dfrac{25}{12}\dfrac{a(x_n)}{\Delta x}}\right)\left[\begin{array}{l} -4(y_{n-1} - Y_{n-1}) + 3(y_{n-2} - Y_{n-2}) \\ -\dfrac{4}{3}(y_{n-2} - Y_{n-2}) + \dfrac{1}{4}(y_{n-4} - Y_{n-4}) \\ -\dfrac{1}{5}(\Delta x)^5\left(\dfrac{d^5Y}{dx^5}\right)_{n-1} + \ldots \end{array}\right]. \quad (23)$$

Thus by taking higher order polynomials to fit y at a large number of points, it is possible to obtain progressively more excellent fits of Y. In all cases, the use of forward interpolation in the evaluation of $(dy/dx)_n$ in terms of y_n, y_{n-1}, ... is the mathematical operation which forces y_n to seek out and approximate Y_n.

Acknowledgments.—The authors wish to thank Creighton Buck for suggesting the geometrical interpretation of the forward interpolation scheme; John Tukey for giving the authors the benefit of his experience with the solution of "stiff" equations; Albert Schaeffer, Kenneth Arnold and J. Barkley Rosser for their interest in this problem.

* This work was carried out under Contract NOrd-9938 with the Navy Bureau of Ordnance.

[1] Assuming the value of $a(x, y)$ to be independent of y.

---

*THE ELEMENTARY BASIC PRINCIPLES OF THE UNIFIED THEORY OF RELATIVITY**

BY VÁCLAV HLAVATÝ

DEPARTMENT OF MATHEMATICS, INDIANA UNIVERSITY

Communicated by T. Y. Thomas, January 14, 1952

1. Introduction.—The unified theory of relativity exposed recently by Einstein[1] is based on three principles: (A) The introduction of a non-symmetric tensor $g_{\lambda\mu}$ in the space X_4 of the relativity; (B) the introduction of a non-symmetric connection $\Gamma^{\nu}_{\lambda\mu}$ by means of

$$\frac{\partial}{\partial x^{\lambda}} g_{\omega\mu} = \Gamma^{\nu}_{\omega\lambda} g_{\nu\mu} + \Gamma^{\nu}_{\lambda\mu} g_{\omega\nu} \tag{1}$$

and finally (C) the introduction of a (seemingly) overdetermined system of conditions imposed on the $\Gamma^{\nu}_{\lambda\mu}$, which yields $g_{\lambda\mu}$.

In the subsequent sections we will deal with each of these basic principles. However, we confine ourselves to results only. The corresponding detailed proofs will be given in a subsequent series of three papers in the *Journal of Rational Mechanics and Analysis.*

2. Principle A.—Denote by $h_{\lambda\mu}(k_{\lambda\mu})$ the symmetric (the skew symmetric) part of $g_{\lambda\mu}$ and by g, h, k the corresponding determinants. Throughout this paper we assume $h \neq 0$. If $n = 4$ and if $h_{\lambda\mu}$ is of the signature $+++-$ then there are in general two sets of bivectors B_1, B_2 totally perpendicular, which are privileged in the sense that they are polar conjugate with respect to the cone $h_{\lambda\mu}$ as well as with respect to the linear complex $k_{\lambda\mu}$ of bivectors. Closely connected with them are four sets of (imaginary) bivectors each of them being polar self conjugate with respect to $h_{\lambda\mu}$ and $k_{\lambda\mu}$.

Projecting this configuration from an arbitrary point P of our space X_4 into the ideal space H_3 of the tangent space $T_4(P)$ of X_4 at P, one obtains a linear line complex K containing a linear line congruence C (whose axes

are the ideal lines of $B_1(P)$ and $B_2(P)$). Both line manifolds K and C are defined intrinsically, by means of $g_{\lambda\mu}$ only.

In another paper[2] I have shown that the projective three-dimensional spinor space, S_3, associated with the (centered) four-dimensional vector space of the (special) theory of relativity is fully characterized by a linear line complex and a linear line congruence contained in it. Also these manifolds are defined intrinsically. Identifying these manifolds with K and C we may identify H_3 and S_3. This identification is not affected by the coördinate transformations (which are different in H_3 and S_3) because in both cases the line manifolds are defined intrinsically. Hence *the introduction of a non-symmetric tensor $g_{\lambda\mu}$ in X_4 leads in the most natural way to the theory of the spinor space of the relativity theory.*

3. Principle B.—In the subsequent sections we denote by $\left\{ \begin{matrix} \nu \\ \lambda\mu \end{matrix} \right\}$ the Christoffel symbols of $h_{\lambda\mu}$, by $H_{\omega\mu\lambda}^{\nu}$ its curvature tensor, by ∇_μ the symbol of covariant derivative with respect to $\left\{ \begin{matrix} \nu \\ \lambda\mu \end{matrix} \right\}$. Moreover we put $\Delta = h^{\lambda\mu}\nabla_\lambda\nabla_\mu$ and $^\circ\Delta = {}^\circ h^{\lambda\mu}\dfrac{\partial^2}{\partial x^\lambda \partial x^\mu}$, where ${}^\circ h^{\lambda\mu} = h^{\lambda\mu}(P)$. The curvature tensor of $\Gamma_{\lambda\mu}^{\nu}$ will be denoted by $P_{\omega\mu\lambda}^{\nu}$. If $n = 4$ then $e^{\omega\mu\lambda\nu}$ denotes the four vector Ricci density (with components $+1$, -1, 0).

The equations (1) split in two systems

$$(a) \qquad \Gamma_{\lambda\mu}^{\nu} = \left\{ \begin{matrix} \nu \\ \lambda\mu \end{matrix} \right\} + S_{\lambda\mu}^{\nu} - h^{\alpha\nu}(S_{\alpha\lambda}^{\gamma}k_{\gamma\mu} + S_{\alpha\mu}^{\gamma}k_{\gamma\lambda})$$

$$(b) \qquad \nabla_\omega k_{\lambda\mu} = 2S_{\alpha\beta}^{\gamma}X_{\omega\mu\lambda\gamma}^{\alpha\beta} \qquad (2)$$

where $X_{\omega\mu\lambda\gamma}^{\alpha\beta}$ is a tensor, function of $h_{\lambda\mu}$ and $k_{\lambda\mu}$. Computing $S_{\alpha\beta}^{\gamma}$ from $(2(b))$ and substituting it in $(2(a))$ one obtains the solution $\Gamma_{\lambda\mu}^{\nu}$ of (1). A necessary condition for $(2(b))$ to admit exactly one solution is $g \neq 0$. A sufficient condition for $(2(b))$ to admit more than one solution $S_{\alpha\beta}^{\gamma}$ is $g = 0$. However, there are cases where $(2(b))$ does not admit any solution $S_{\alpha\beta}^{\gamma}$ (and consequently (1) does not admit any solution $\Gamma_{\lambda\mu}^{\nu}$).

4. Almost Riemannian Space.—Our space X_n will be termed an almost Riemannian space R_n if $k_{\lambda\mu} = \epsilon K_{\lambda\mu}$, $S_{\lambda\mu}^{\nu} = \epsilon s_{\lambda\mu}^{\nu}$, where $\epsilon \to 0$ is a constant and $K_{\lambda\mu}$, $s_{\lambda\mu}^{\nu}$ are tensors different from zero. In R_n we obtain from $(2(b))$[3]

$$(a) \qquad s_{\omega\mu}^{\nu} \equiv \frac{1}{2}h^{\nu\alpha}(\nabla_\alpha K_{\omega\mu} + \nabla_\omega K_{\alpha\mu} - \nabla_\mu K_{\alpha\omega}) \pmod{\epsilon} \qquad (3)$$

and

$$(b) \qquad \Gamma_{\lambda\mu}^{\nu} \equiv \left\{ \begin{matrix} \nu \\ \lambda\mu \end{matrix} \right\} + \epsilon\,[s_{\lambda\mu}^{\nu} - h^{\nu\alpha}(s_{\alpha\lambda}^{\gamma}k_{\gamma\mu} + s_{\alpha\mu}^{\gamma}k_{\gamma\lambda})] \pmod{\epsilon^2}. \qquad (3)$$

Substituting from (3(*a*)) in (3(*b*)) we obtain the solution $\Gamma^{\nu}_{\lambda\mu}$ of (1).

5. *Principle C.*—If $n = 4$ the conditions

$$S^{\mu}_{\lambda\mu} = 0 \qquad P^{\nu}_{\nu[\mu\lambda]} = 0 \qquad e^{\omega\mu\lambda[\nu}P^{\rho]}_{\omega\mu\lambda} = 0 \tag{4}$$

represent $4 + 6 + 6 = 16$ differential equations for 16 unknowns $h_{\lambda\mu}$ and $k_{\lambda\mu}$. These conditions reduce in R_4 to

$$(a) \qquad \nabla^{\lambda}K_{\lambda\mu} \equiv 0 \pmod{\epsilon}, \qquad (\nabla^{\lambda} = h^{\lambda\mu}\nabla_{\mu})$$

$$(b) \qquad \Delta K_{\lambda\mu} + h^{\alpha\beta}(H^{\rho}_{\lambda\mu\alpha}K_{\rho\beta} + 2H_{\rho\alpha[\lambda}{}^{\rho}K_{\mu]\beta}) \equiv 0 \pmod{\epsilon}$$

$$(c) \qquad \nabla^{[\nu}I^{\rho]} + 2e^{\omega\mu\lambda[\rho}(h^{\nu]\alpha}H^{\beta}_{\alpha\lambda\omega}K_{\mu\beta} - K_{\alpha\mu}H^{\nu]}_{\lambda\omega\beta}h^{\alpha\beta}) \equiv 0 \pmod{\epsilon} \tag{5}$$

where $I^{\rho} = e^{\omega\mu\lambda\rho}\nabla_{\omega}K_{\mu\lambda}$ is (up to a numerical factor) the Einstein electric current vector density. In order to find a solution of (5) we introduce a neighborhood $N_{\epsilon}(P)$ of a general point $P(\overset{\circ}{x}): N_{\epsilon}(P)$ is the set of all points whose coördinates x^{ν} satisfy the condition $|(x^{\nu} - \overset{\circ}{x}{}^{\nu})^3| \leq \epsilon$. If

$$h_{\lambda\mu} = {}^{\circ}h_{\lambda\mu} + \epsilon\, h'_{\lambda\mu} \text{ in } N_{\epsilon}(P) \tag{6}$$

then (5) reduces to

$${}^{\circ}\Delta K_{\lambda\mu} \equiv 0 \pmod{\epsilon}, \qquad \frac{\partial}{\partial x}{}^{[\lambda}I^{\mu]} \equiv 0 \pmod{\epsilon}.$$

It is not difficult to find a solution $K_{\lambda\mu}$ of this system. Thus for instance

$$K_{12} \equiv A(z - t),\ K_{34} \equiv B(x + iy),\ K_{\lambda\mu} \equiv 0 \text{ for } \lambda\mu \neq 12, 34 \pmod{\epsilon} \text{ in } N_{\epsilon}(P)$$

is a solution, provided B is an analytic (regular) function of $x + iy$ (and ${}^{\circ}h_{11} = {}^{\circ}h_{22} = {}^{\circ}h_{33} = -{}^{\circ}h_{44} = 1$, the remaining ${}^{\circ}h_{\lambda\mu} = 0$).

6. *The First Einstein Conditions.*—If we put

$$R_{\lambda\mu} = \frac{\partial}{\partial x^{\alpha}}\,\Gamma^{\alpha}_{\lambda\mu} - \Gamma^{\alpha}_{\lambda\beta}\Gamma^{\beta}_{\alpha\mu} - \frac{1}{2}\left(\frac{\partial}{\partial x^{\mu}}\,\Gamma^{\alpha}_{(\lambda\alpha)} + \frac{\partial}{\partial x^{\lambda}}\,\Gamma^{\alpha}_{(\mu\alpha)}\right) + \Gamma^{\alpha}_{\lambda\mu}\Gamma^{\beta}_{(\alpha\beta)}$$

then Einstein's conditions imposed on the $\Gamma^{\nu}_{\lambda\mu}$ are

$$S^{\mu}_{\lambda\mu} = 0, \qquad R_{\lambda\mu} = 0 \tag{7}$$

and represent in this form $n + n^2$ conditions for n^2 unknowns $h_{\lambda\mu}$ and $k_{\lambda\mu}$. However, in R_n they split in $n\dfrac{(n+1)}{2} + n\dfrac{(n-1)}{2} = n^2$ conditions namely

$$(a) \quad H^{\nu}_{\nu\mu\lambda} \equiv 0 \pmod{\epsilon^2}, \qquad (b)\ \Delta K_{\lambda\mu} + h^{\alpha\beta}H^{\rho}_{\lambda\mu\alpha}K_{\rho\beta} \equiv 0 \pmod{\epsilon}. \tag{8}$$

This system admits a solution $h_{\lambda\mu} \equiv \sigma\,{}^{\circ}h_{\lambda\mu} \pmod{\epsilon^2}$ in $N_{\epsilon}(P)$ where σ is an explicitly defined function of position. Moreover we have in this

case $H^{\nu}_{\omega\mu\lambda} \equiv 0 \pmod{\epsilon^2}$ in $N_\epsilon(P)$. Hence a coördinate system may be found for which $h_{\lambda\mu}$ are constant in $N_\epsilon(P)$ so that (8(b)) reduces to ${}^\circ\Delta K_{\lambda\mu} \equiv 0 \pmod{\epsilon}$ in $N_\epsilon(P)$. Moreover one obtains the same condition if (6) holds, or if one assumes that the set of bivectors $B_1(B_2)$ generates a non-holonomic surface which differs only mod ϵ in $N_\epsilon(P)$ from a set of holonomic surfaces. However, in this case (8(a)) leads to $h_{\lambda\mu} \equiv {}^\circ h_{\lambda\mu} \pmod{\epsilon^2}$ in $N_\epsilon(P)$ while if (6) holds the equation (8(a)) reduces to

$$ {}^\circ\Delta h'_{\lambda\mu} \equiv {}^\circ h^{\nu\alpha}\left(\frac{\partial^2}{\partial x^\alpha \partial x^\mu} h'_{\lambda\nu} + \frac{\partial^2}{\partial x^\alpha \partial x^\lambda} h'_{\nu\mu} - \frac{\partial^2}{\partial x^\mu \partial x^\lambda} h'_{\alpha\nu}\right) \pmod{\epsilon}. \quad (9) $$

It is not difficult to find a non-trivial solution of (9). In particular if

$$ {}^\circ h^{\nu\alpha} \frac{\partial}{\partial x^\nu} h'_{\alpha\lambda} \equiv \frac{1}{2} {}^\circ h^{\nu\alpha} \frac{\partial}{\partial x^\lambda} h'_{\nu\alpha} \pmod{\epsilon}^4 $$

then (9) reduces to

$$ {}^\circ\Delta h'_{\lambda\mu} \equiv 0 \pmod{\epsilon} \quad (10) $$

If ${}^\circ h_{11} = {}^\circ h_{22} = {}^\circ h_{33} = -{}^\circ h_{44} = 1$ (and the remaining ${}^\circ h_{\lambda\mu} = 0$), then (10) admits a solution

$$ h'_{\lambda\mu} \equiv p_{\lambda\mu}(x + iy, z - t) \pmod{\epsilon} \text{ in } N_\epsilon(P) $$

where $p_{\lambda\mu} = p_{\mu\lambda}$ are arbitrary (regular) functions of $x + iy$.

7. *The Second Einstein Conditions.*—Any solution of (8) is also a solution of the second Einstein system

$$ S^\mu_{\lambda\mu} = 0 \qquad R_{(\mu\lambda)} = 0 \qquad \frac{\partial}{\partial x^{[\omega}} R_{\mu\lambda]} = 0 \quad (11) $$

which (for $n = 4$) represents $4 + 10 + 4 = 18$ conditions for 16 unknowns. This system reduces in R_4 to

$$ (a) \quad H^{\nu}_{\nu\omega\mu} \equiv 0 \pmod{\epsilon^2} $$

$$ (b) \quad \Delta I^\nu + e^{\omega\mu\lambda\nu}(H^{\rho}_{\lambda\mu\alpha}\nabla_\omega K_{\beta\rho} - 4H^{\rho}_{\alpha\omega\lambda}\nabla_\beta K_{\mu\rho})h^{\alpha\beta} \equiv 0 \pmod{\epsilon}. \quad (12) $$

In all cases mentioned in section (6) we obtain the same solution for $h_{\lambda\mu}$, while (in these cases) the equation (12(b)) reduces to ${}^\circ\Delta I^\nu \equiv 0 \pmod{\epsilon}$. If ${}^\circ h_{11} = {}^\circ h_{22} = {}^\circ h_{33} = -{}^\circ h_{44} = 1$ the remaining ${}^\circ h_{\lambda\mu} = 0$, then $I^\nu \equiv f^\nu(x + iy, z - t) \pmod{\epsilon}$ in $N_\epsilon(P)$ is a solution of this equation, provided that f^ν is analytic (regular) in $x + iy$.

* Prepared under Army Contract DA-33-008-Ord-224

[1] Einstein, A., *The Meaning of Relativity*, Princeton University Press, 1950, Appendix II.

[2] Hlavatý, V., "Spinor Space and Line Geometry," *Canadian J. Math.*, **3**, 442–459 (1950), and "Spinor Space and Line Geometry II" (to be published in the *Journal of*

Rational Mechanics and Analysis) where the Dirac equations are interpreted by means of the complex and the congruence mentioned above.

[3] $A \equiv B \pmod{\epsilon^m}$ is the analytical description of the statement that the difference $A - B$ is of the magnitude order of $\epsilon^{m'}$, $m' \geq m$.

[4] It is worth while to stress that (9) and (10) are covariant equations [thanks to (6)]. In particular (10) cannot be enforced by a proper choice of coördinates!

HOMOLOGY RELATIONS ON REGULAR ORIENTABLE MANIFOLDS

BY MARSTON MORSE

INSTITUTE FOR ADVANCED STUDY

Communicated January 25, 1952

1. Introduction.—The theorems stated in 1927 by Morse in reference 1 are becoming increasingly significant in questions of homology, homotopy and category. Studies related to the unsolved Schoenflies problem by Morse and Baiada are based on these theorems. It is for this reason that the unpublished proofs of these theorems as written in 1927 will be here given with relatively obvious extensions whereby chains over the field of integers mod 2 are replaced by chains over any field Φ. The connectivities used in reference 1 will here be replaced by the dimensions R_i of the respective homology groups, always using cycles over Φ.

The complexes involved will be similar to those used in reference 2. They are broken up into cells with the properties of "Block-ketten" of reference 5, page 78. The ranks of the incidence matrices will be determined "relative" to Φ, replacing the elements a of a matrix by ea, where e is the unit in Φ, and evaluating the subdeterminants of the resultant matrix as elements in Φ.

2. Regular Manifolds and Regions.—A regular n-manifold $\sum_n$ in a Euclidean $(n + r)$-space E_{n+r}, $n > 0$, $r > 0$ is a closed, compact, connected Hausdorff subspace of E_{n+r} specially representable as follows. Let U be an arbitrary open set in a Euclidean space E_n with rectangular coordinates $u = (u_1, \ldots, u_n)$. Let $y = (y_1, \ldots, y_{n+r})$ similarly represent a point in E_{n+r}. Let P be an arbitrary point of $\sum_n$. Some neighborhood $N(P)$ of P relative to $\sum_n$ shall be the homeomorph of an open set $U \in E_n$ under a mapping of $u \in U$ onto $[\psi_1(u), \ldots, \psi_{n+r}(u)] \in E_{n+r}$ where the functions ψ_i are of class C^m, $m > 0$, and possess a functional matrix of rank n. By a *regular subregion* of $\sum_n$ will be meant a closed non-empty subset of $\sum_n$ bounded by a finite set of disjoint regular $(n - 1)$-manifolds. Unless otherwise stated all manifolds including boundaries of regions are to be regular and *given* by representations of class C'''. Other regular representations of a lower class may sometimes be used.

By means of relatively simple homeomorphic approximations it is possible to establish our final theorems for manifolds and regions of class C' and even for polyhedral manifolds and regions. We shall not becloud the principal problems with such details. Another trivial extension is to coordinate manifolds and regions. We have chosen the Euclidean case for simplicity of statement of hypotheses.

Because of local connectedness of our manifolds and regions it is immaterial whether homologies are taken in the sense of the singular theory or Vietoris theory. The basic determination of the R_i is, however, combinatorial by means of matrices of incidence as in reference 2.

3. Non-degenerate Functions f.—We shall consider real functions f with values $f(y)$ defined for $y \in \sum_n$. If ψ: $U \longleftrightarrow N(P)$ is regular and of class C''' over U, $f(\psi)$ shall be given as of at least class C'' over U for each $P \in \sum_n$. A critical point (u_0) of $f(\psi)$ defines a *critical point* $\psi(u_0)$ of f and is termed *non-degenerate* if the Hessian of $f(\psi)$ does not vanish at (u_0). When (u_0) is a critical point of $f(\psi)$, $f(\psi)$ shall be given as of class C''' over some sufficiently restricted neighborhood of (u_0). Given a critical point P_0 of f and critical value c of f it follows as in reference 4, page 173, that there exists a mapping φ: $U \longleftrightarrow N(P_0)$ regular and of class C' such that $f(\varphi) - c$ has the values over U of a form quadratic in the coordinates of U. Such a representation of $f|N(P_0)$ will be termed *canonical*.

By methods exemplified in reference 4, page 243, it is easy to establish the existence of non-degenerate functions f over $\sum_n$. By the methods of reference 3 one can modify the values of f on the boundaries of regular regions on which f has no critical point so as to make the gradient of f constantly entrant or emergent, and f a constant on each boundary manifold.

Let f_a and f^a be, respectively, the sets $\sum_n | f(y) \leqq a$, and $\sum_n | f(y) = a$. If $\alpha \leqq \beta$ let $f_{\alpha\beta}$ be the set $\sum_n | \alpha \leqq f(y) \leqq \beta$. Our ultimate theorems are independent of f and concern regular regions C_n, their boundaries BC_n and their complements on $\sum_n$. By suitable choice of non-degenerate functions f, in accordance with the existence theorems of the preceding paragraph, one can reduce our theorems on regular regions and their complements to theorems on $f_{\alpha\beta}$ and $f_{\beta\gamma}$ where $\alpha < \min f < \beta < \max f < \gamma$. When a given function is modified its class C^n will in general be lowered. In all cases it will be of at least class C'.

4. Variation of $R_i(f_a)$ with a.—This variation is analyzed in reference 2 in the case of a regular subregion of a Euclidean space. Reference 2, with trivial modifications, thus affords a local analysis of f when locally represented in the form $f(\psi)$ in terms of the Euclidean coordinates of U. Reference 2 can be simplified by the deletion of sections 10 and 11, since these sections become trivial when f is canonically represented. The use

of orthogonal trajectories of level manifolds of f naturally takes the differential form appropriate to a curved space $\sum_n$, see reference 4, page 152. Sections 13, 14, 15 of reference 2 can be taken over here with the following modifications. We suppose all cells oriented; a corresponding modification of the matrices of incidence is required. In particular the $k-1$ matrices of incidence of the $(k-1)$-sphere S_{k-1} shall have two columns, each consisting of a 1 above a -1.

The chains in reference 2 were functions with values over the field of integers mod 2. Here the field Φ is general and the present connectivity R_i, defined as the dimension of the ith homology group, accordingly differs from R_i in reference 2. An additional difference arises in that even when Φ is the field of integers mod 2, and $R_i(C_n)$ is a connectivity in reference 2, our present connectivity, using the same field, is $R_i(C_n) - 1$ when $i = 1, \ldots, n$. For an n-sphere, for example, we here have $R_0 = R_n = 1$, $R_1 = R_2 = \ldots = R_{n-1} = 0$. Even with these changes in notation equations (134) of reference 2, giving the differences ΔR_i taken in the present sense, are applicable, provided the ranks ρ_i of the matrices of incidence are appropriately taken relative to Φ.

As in reference 2 consider a critical point P_0 of index k with $0 < k < n$. We can suppose that the critical value $f(P_0) = c$ is assumed by no other critical point with f-value on the interval $(c-\epsilon, c+\epsilon)$. In reference 2, $\epsilon = e^2$. As in reference 2, §15, let C_n be the complex of unoriented cells decomposing $f_{c+\epsilon}$ so as to include the "new cell" $(a_k b_{n-k})$and its boundary cells, and let C'_n be C_n with the cell $(a_k b_{n-k})$ and the new boundary cells deleted. Let the cells of C_n and C'_n now be oriented and the resulting complexes denoted by Γ_n and Γ'_n, respectively. The special product cells $(a_k b_{n-k})$, $(a'_{k-1} b'_0)$, etc., of reference 2 will thereby be oriented and denoted by $[a_k b_{n-k}]$, $[a'_{k-1} b'_0]$, etc.

Let J_i and J'_i, $i = 1, \ldots, n$, be the matrices of incidence of Γ_n and Γ'_n, respectively. The matrix J_k will have the form of the matrix H_k on page 383 of reference 2, provided the submatrix H'_k is replaced by J'_k and the last two columns are appropriately modified. These columns are to be headed by $[a_k b'_0]$ and $[a_k b''_0]$ and their only non-vanishing elements will be those in the submatrix

	$[a_k b'_0]$	$[a_k b''_0]$
$[a'_{k-1} b'_0]$	1	0
$[a''_{k-1} b'_0]$	-1	0
$[a'_{k-1} b''_0]$	0	1
$[a''_{k-1} b''_0]$	0	-1

Let ρ_i and ρ'_i be the ranks relative to Φ of J_i and J'_i, respectively. Operating on the matrix J_k essentially as on H_k in reference 2, taking account of

Φ, one finds that two cases may occur. In Case 1, $\rho_k = \rho'_k$, and in Case 2, $\rho_k = \rho'_k + 1$.

Our present matrix J_{k+j}, $j = 1, 2, \ldots, n - k - 1$, similarly has the form of H_{k+j} as represented on page 384 of reference 2, provided one replaces H'_{k+j} by J'_{k+j} and appropriately modifies the last two columns. The last two columns are to be headed by $[a_k b'_j][a_k b''_j]$ and the four 1's in each of these columns in H_{k+j} here replaced by $1, -1, 1, -1$, respectively. One similarly obtains J_n from H_n of reference 2 by replacing H'_n by J'_n and modifying the last column of H_n.

As in reference 2, one infers that

$$\rho_{k+j} = \rho'_{k+j} + 1 \qquad (j = 1, 2, \ldots, n - k - 1)$$

and $\rho_n = \rho'_n + 1$. The equations at the bottom of page 384 of reference 2 can then be applied to Γ_n and Γ'_n. Hence, as a increases through a critical value c assumed by a critical point P_0 of index k, $0 < k < n$, either $R_k(f_a)$ increases by 1 or else $R_{k-1}(f_a)$ decreases by 1. The numbers $R_i(f_a)$ otherwise remain constant.

5. Variation of $R_i(f^a)$ with a.—We continue with the critical point P_0, the critical value c and the constant ϵ. From the orientability of $\sum_n$ it follows that for a an ordinary value of f the $(n - 1)$-manifold f^a is orientable and hence the Poincaré law $R_i(f^a) = R_{n-i-1}(f^a)$ holds. The change in $R_i(f^a)$ as a increases through c is sought. The alternatives depend upon the value of the index k and upon n.

Let $B\Gamma_n$ be the $(n - 1)$-complex of oriented cells on the geometric boundary of Γ_n. The carrier $|B\Gamma_n|$ of $B\Gamma_n$ is $f^{c+\epsilon}$. Although $|B\Gamma'_n| \neq f^{c-\epsilon}$ it is readily seen that $|B\Gamma'_n|$ is the homeomorph of $f^{c-\epsilon}$. We introduce the $(n - 1)$-complex

$$X_{n-1} = B\Gamma_n \cup B\Gamma'_n, \tag{5.0}$$

taking the cells of X_{n-1} with the orientation and incidence relations they have in Γ_n. One obtains X_{n-1} from $B\Gamma'_n$ by adding the "new-cells" on $B[a_k b_{n-k}]$. Let K_i and K'_i be, respectively, the ith incidence matrices of X_{n-1} and $B\Gamma'_n$ $(i = 1, \ldots, n - 1)$. Suppose $0 < k < n$. For $i = 1, 2, \ldots, k - 1$, $K_i = K'_i$. The matrix K_k has the form of the matrix H_k on page 383 of reference 2, provided K'_k replaces H'_k therein, and the last two columns of H_k are modified as in section 4. The matrix K_{k+j}, $j = 1, 2, \ldots, n - k - 1$, has the form of the matrix H_{k+j} at the top of page 384, provided K'_{k+j} replaces H'_{k+j} in H_{k+j} and the last two columns of H_{k+j} are modified as in section 4. The equations at the bottom of page 384 then hold for $i = 1, 2, \ldots, n - 1$, where α_i and ρ_i refer to K_i, and α'_i and ρ'_i refer to K'_i. Equations (134) of reference 2 are here replaced by the equations, $i = 0, 1, \ldots, n - 1$,

$$R_i(X_{n-1}) - R_i(B\Gamma'_n) = (\alpha_i - \alpha'_i) - (\rho_i - \rho'_i) - (\rho_{i+1} - \rho'_{i+1}), \quad (5.1)$$

where $\rho_0 = \rho'_0 = \rho_n = \rho'_n = 0$. Recalling that $|B\Gamma'_n|$ is the homeomorph of $f^{c-\epsilon}$ we have the result

(*a*) *When $0 < k < n$ the connectivities of X_{n-1} and $f^{c-\epsilon}$ are equal except for the mutually exclusive changes*

$$\left.\begin{aligned} R_k(X_{n-1}) &= R_k(f^{c-\epsilon}) + 1 \quad \text{Case 1a,} \\ R_{k-1}(X_{n-1}) &= R_{k-1}(f^{c-\epsilon}) - 1 \quad \text{Case 2a,} \end{aligned}\right\} \quad (5.2)$$

and except for an additional change, independent of k,

$$R_{n-1}(X_{n-1}) = R_{n-1}(f^{c-\epsilon}) + 1. \quad (5.3)$$

The complex X_{n-1} mediates between $B\Gamma'_n$ and $B\Gamma_n$, or if one pleases between $f^{c-\epsilon}$ and $f^{c+\epsilon}$. In this sense (*b*) completes (*a*). The change $\Delta R_{n-1} = 1$ due to (5.3) is to be compounded with the change in R_{n-1} under Case 1a or 2a when $k = n - 1$, making a resultant $\Delta R_{n-1} = 2$ or 1 in Case 1a or Case 2a, respectively. Similarly with (5.4) in (*b*).

(*b*) *When $0 < k < n$ the connectivities of X_{n-1} and $f^{c+\epsilon}$ are equal except for the mutually exclusive changes*

$$\begin{aligned} R_{n-k-1}(X_{n-1}) &= R_{n-k-1}(f^{c-\epsilon}) - 1 \quad \text{Case 1b} \\ R_{n-k}(X_{n-1}) &= R_{n-k}(f^{c+\epsilon}) + 1 \quad \text{Case 2b} \end{aligned}$$

and except for an additional change, independent of k,

$$R_{n-1}(X_{n-1}) = R_{n-1}(f^{c+\epsilon}) - 1. \quad (5.4)$$

Observe that X_{n-1} can be obtained from $B\Gamma_n$ by adding the "old cells" on $B[a_k b_{n-k}]$ not already in $B\Gamma_n$. Let K''_i be the ith matrix of incidence of $B\Gamma_n$, $i = 1, 2, \ldots, n-1$. Then $K_i = K''_i$ for $i = 1, 2, \ldots, n-k-1$. The matrix K_{n-k} has the form of the matrix on page 383, provided H'_k therein is replaced by K''_{n-k} and provided the non-zero elements in the last two columns are those of the submatrix

	$[a'_0 b_{n-k}]$	$[a''_0 b_{n-k}]$
$[a'_0 b'_{n-k-1}]$	1	0
$[a'_0 b''_{n-k-1}]$	−1	0
$[a''_0 b'_{n-k-1}]$	0	1
$[a''_0 b''_{n-k-1}]$	0	−1

The matrix K_{n-k+j}, $j = 1, 2, \ldots, k-1$, similarly has the form of the matrix at the top of page 384 provided K''_{n-k+j} replaces H'_{k+j} therein, the last two columns are headed by $[a'_j b_{n-k}]$ and $[a''_j b_{n-k}]$, the last six rows headed by $[a'_j b'_{n-k-1}]$ $[a'_j b''_{n-k-1}]$, etc., while the four 1's in each of the last two columns are replaced by 1, −1, 1, −1, respectively. If ρ_i and

ρ'_i are, respectively, the ranks of X_{n-1} and $B\Gamma_n$ relative to Φ, then either $\rho_{n-k} = \rho'_{n-k}$ or $\rho_{n-k} = \rho'_{n-k} + 1$ while $\rho_{n-k+j} = \rho'_{n-k+j} + 1$, $j = 1, 2, \ldots, k-1$. If α_i and α'_i in the equations at the bottom of page 384 then refer to X_{n-1} and $B\Gamma_n$, respectively, these equations hold for $i = 1, 2, \ldots, n-1$. Statement (b) follows.

Statements (a) and (b) imply the following theorem.

THEOREM 5.1. *If a value c of f is assumed at just one non-degenerate critical point of index k then the differences $\Delta B_i = R_i(f^{c+\epsilon}) - R_i(f^{c-\epsilon})$ for a sufficiently small $\epsilon > 0$ have the following values:*

$$\Delta B_0 = \Delta B_{n-1} = 1 \qquad (k = 0) \tag{5.5}$$

$$\Delta B_{n-1} = \Delta B_0 = -1 \qquad (k = n) \tag{5.6}$$

$$\begin{array}{lll} \text{Either} & \Delta B_k = \Delta B_{n-k-1} = 1 & \\ \text{or} & \Delta B_{k-1} = \Delta B_{n-k} = -1 & \end{array} \qquad \begin{array}{l} k = 1, 2, \ldots, n-1 \\ 2k \neq n-1, n, n+1 \end{array} \tag{5.7}$$

$$\begin{array}{lll} \text{Either} & \Delta B_k = \Delta B_k = 2 & \\ \text{or} & \Delta B_{k-1} = \Delta B_{k+1} = -1 & \end{array} \qquad (2k = n-1) \tag{5.8}$$

$$\begin{array}{lll} \text{Either} & \Delta B_k = \Delta B_{k-1} = 1 & \\ \text{or} & \Delta B_{k-1} = \Delta B_k = -1 & \\ \text{or} & \Delta B_{k-1} = \Delta B_k = 0 & \end{array} \qquad (2k = n) \tag{5.9}$$

$$\begin{array}{lll} \text{Either} & \Delta B_k = \Delta B_{k-2} = 1 & \\ \text{or} & \Delta B_{k-1} = \Delta B_{k-1} = -2 & \end{array} \qquad (2k = n+1) \tag{5.10}$$

All Other Differences $\Delta B_i = 0$:

In proving Theorem 5.1 one passes from $f^{c-\epsilon}$ to X_{n-1} to $f^{c+\epsilon}$ using (a) and (b). In deriving the values of ΔB_i resulting from the changes implied by (a) and (b), equations (5.3) and (5.4) can be neglected since their resultant effect on ΔB_{n-1} is zero. Apart from equations (5.3) and (5.4), four algebraically different indices appear in (a) and (b). The conditions that two of these indices have equal values give rise to side conditions in Theorem 5.1. In particular the equality $k = n - k - 1$ implies $2k = n - 1$, $k - 1 = n - k$ implies $2k = n + 1$, and the two remaining equalities, $k = n - k$ and $k - 1 = n - k - 1$, both imply $2k = n$. Let two indices i and j on the interval $[0, n-1]$ be termed *dual* if $i + j = n - 1$. The conditions $2k = n - 1$, and $2k = n + 1$ imply that k and $n - k$ are self dual, respectively, while $2k = n$ implies that k and $k - 1$ are dual. The two indices appearing in each equality in (5.1) are dual. According to Poincaré, $\Delta B_i = \Delta B_j$ if i and j are dual.

The cases $k = 0$ or n are trivial. When $0 < k < n$ Cases 1a or 2a may be combined with Cases 1b or 2b giving four *a priori* combinations. Apart from the special case $2k = n$, the Poincaré law, however, implies that

Cases 1a and 1b occur together and Cases 2a and 2b likewise, as we shall see.

Consider equations (5.7). In Case 1a, $\Delta R_k = 1$ as one changes from $f^{c-\epsilon}$ to X_{n-1}, neglecting (5.3) when $k = n - 1$. Under the side condition $2k \neq n - 1$, n of (5.7) one has $k \neq n - k - 1$, $n - k$, so that on passing from X_{n-1} to $f^{c-\epsilon}$ in accordance with (b) $\Delta R_k = 0$, neglecting (5.4) when $k = n - 1$. Thus (a) and (b) imply a resultant change, $\Delta B_k = 1$, on passing from $f^{c-\epsilon}$ to $f^{c+\epsilon}$; by virtue of the Poincaré law one has a resultant $\Delta B_{n-k-1} = \Delta B_k = 1$. As a consequence, Case 1b must hold with Case 1a. Similarly, in Case 2a, $\Delta B_{k-1} = \Delta B_{n-k} = -1$ on passing from $f^{c-\epsilon}$ to $f^{c+\epsilon}$, so that Case 2b must hold.

The remaining equations similarly follow. The first alternative in (5.8), (5.9), (5.10) occurs when Cases 1a and 1b both occur, the second when Cases 2a and 2b both occur. When $2k = n$ the third alternative in (5.9) occurs when Cases 1a and 2b both occur, or Cases 2a and 1b.

6. Critical Points of Increasing, Decreasing and Neutral Types.—Let Z_a be any one of the subsets f_a, f^a, Comp (f_a), f_{aa}, $f_{a\beta}$. As a increases through a value c assumed at just one non-degenerate critical point P_0 the changes $\Delta R_i(Z_a)$ are of one sign or zero, independent of i. If each such difference vanishes P_0 will be said to be of *neutral* type rel Z_a. If at least one such difference is positive (negative) P_0 will be said to be of *increasing* (*decreasing*) type rel Z_a.

THEOREM 6.1. *If P_0 is of decreasing type rel f_a, it will be of decreasing type rel f^a.*

Let k be the index of P_0.

Case I: $k = 0$ or n. When $k = 0$, P_0 is always of increasing type rel f_a, and when $k = n$, always of decreasing type rel f^a. The theorem follows.

Case II: $2k \neq n$, $0 < k < n$. As has been seen in the analysis of the matrix of incidence J_k of Γ_n in section 4, P_0 is of decreasing type rel f_a if and only if $\rho_k = \rho'_k + 1$, where ρ_k and ρ'_k are, respectively, the ranks of J_k and J'_k, or equivalently if the cycle boundary $\beta[a_k b'_0]$ of $[a_k b'_0]$ does not bound on Γ'_n. Turning to f^a recall that when $2k \neq n$, P_0 is of decreasing type rel f^a if and only if Case 2a occurs in (a) in section 5, or equivalently if $\rho_k = \rho'_k + 1$ where ρ_k and ρ'_k are the ranks of K_k and K'_k, respectively, or equivalently if and only if $\beta[a_k b'_0]$ not ~ 0 on $B\Gamma'_n$. The latter condition is implied by $\beta[a_k b'_0]$ not ~ 0 on Γ'_n. The theorem follows in Case II.

Case III: $2k = n$, $0 < k < n$. By virtue of Case II the theorem holds if n is odd. We take advantage of this fact when $n = 2k$ by replacing $\sum_n$ by a manifold $\sum_{n+1}$ of odd dimension and f by a function F appropriately defined on $\sum_{n+1}$.

The Definition of $\sum_{n+1}$ and F: We are supposing that $\sum_n$ lies in a space E_{n+r} of points y with coordinates $(y_1, \ldots, y_{n+r})$. Without loss of generality

one can suppose that $y_1 \mid \sum_n \neq 0$. Let E_{n+r} be regarded as the subspace of the space E_{n+r+1} of points y' and coordinates $(y_0, y_1, \ldots, y_{n+r})$ on which $y_0 = 0$. If A is any subset of E_{n+r} on which $y_1 > 0$ let ΘA be a subset of E_{n+r+1} generated by revolving A around E_{n+r} as follows. For $y = (y_1, \ldots, y_{n+r}) \in A$ let Θy be the circle of points in E_{n+k+1} with coordinates

$$(y_1 \cos\theta, y_1 \sin\theta, y_2, \ldots, y_{n+e}) \qquad (0 \leqq \theta < 2\pi), \tag{6.1a}$$

and let ΘA be the union of the circles Θy for $y \in A$. In particular let $\sum_{n+1} = \Theta \sum_n$. One sees that $\sum_{n+1}$ is regular with $\sum_n$. We now define F over $\sum_{n+1}$ by setting

$$Fy_1 \cos\theta, y_1 \sin\theta, (y_2, \ldots, y_{n+r}) = f(y) + \eta \cos\theta \tag{6.1b}$$

for every point $y = (y_1, \ldots, y_{n+r}) \in \sum_n$, where $\eta > 0$ is a constant yet to be chosen.

Observe that F has a critical point on the circle Θy if and only if y is a critical point of f, and if $\theta = 0$ or π. If y has the index k as a critical point of f then a point on the circle Θy has the index $k + 1$, or k, as a critical point of F, according as $\theta = 0$ or π.

Proof of Theorem 6.1 in Case III: Suppose for simplicity that the given critical point P_0 of f has the f-value 0. Let e be so small a positive constant that $[-e, e]$ is an interval of ordinary values of f, 0 excepted. Let $F^\circ = F \mid (\eta = 0)$. The set equalities are clear:

$$F^\circ_{-e} = \Theta f^{-e}, \qquad F^\circ_e = \Theta f^e. \tag{6.1c}$$

For e fixed an obvious use of the orthogonal trajectories of F shows that when $\eta > 0$ is sufficiently small the relations (6.1c) are replaced by the topological equivalences:

$$F^{-e} \longleftrightarrow F^\circ_{-e} = \Theta f^{-e}, \qquad F^e \longleftrightarrow F^\circ_e = \Theta f^e \tag{6.1d}$$

including the equivalences:

$$F^{-e} \longleftrightarrow \Theta f^{-e}, \qquad F^e \longleftrightarrow \Theta f^e. \tag{6.1e}$$

If P_0 is a critical point of f of decreasing type rel f_a, then

$$R_{k-1}(f^e) - R_{k-1}(f^{-e}) = -1. \tag{6.2}$$

In accordance with well-known formulas for product spaces, (6.1d) and (6.2) imply that

$$R_{k-1}(F^e) - R_{k-1}(F^{-e}) = -1, \qquad R_k(F^e) - R_k(F^{-e}) = -1. \tag{6.3}$$

This is possible only if the two critical points P_1 and P_2 of F, with F-values on the interval $(-e, e)$, are both of decreasing type rel F_a. Theorem 6.1 has already been seen to be valid for F, as defined over $\sum_{n+1}$, when the dimension of $\sum_{n+1}$ is odd. In the case at hand $n + 1 = 2k + 1$. By

virtue then of Theorem 6.1, P_1 and P_2, as critical points of F, are of decreasing type rel F^a. Hence by definition

$$R_i(F_c) \leqq R_i(F^{-c}) \qquad (i = 0, \ldots, n), \tag{6.4}$$

where the sign $<$ prevails for some i. Consistent with the product homology theory and (6.1e), with (6.4), and with Theorem 6.1, one can only have

$$R_i(f^c) \leqq R_i f^{-c} \qquad (i = 0, \ldots, n-1), \tag{6.5}$$

where the sign $<$ prevails for some values of i (in particular for $i = k$ and $k-1$ according to (5.9)). Thus P_0 is of decreasing type rel f^a and the theorem holds in Case III.

THEOREM 6.2. *If the critical point P_0 is of decreasing type rel f_a it is of decreasing type rel Comp f_a.*

If one sets $F = -f$ over $\sum_n$, P_0 is a critical point of F of index $n-k$. Moreover for each ordinary value a of f, $R_i(F_{-a}) = R_i(\text{Comp } f_a)$, $i = 0, 1, \ldots, n$. An equivalent way of stating Theorem 6.2 is accordingly as follows. If the critical point P_0 has an index k, and if for every sufficiently small $\epsilon > 0$

$$R_{k-1}(f_{c+\epsilon}) - R_{k-1}(f_{c-\epsilon}) = -1, \tag{6.6}$$

then for every sufficiently small $\epsilon > 0$

$$R_{n-k}(\text{Comp } f_{c+\epsilon}) - R_{n-k}(\text{Comp } f_{c-\epsilon}) = -1. \tag{6.7}$$

To prove Theorem 6.2 suppose P_0 of decreasing type rel f_a. Then P_0 is of decreasing type rel f^a by Theorem 6.1, hence of increasing type rel F^a since $F = -f$, of increasing type rel F_a by Theorem 6.1, and hence of decreasing type rel Comp f_a. This establishes the theorem.

THEOREM 6.3. *If $\alpha < f(P_0)$ and if the critical point P_0 is of decreasing type rel $f_{\alpha a}$, P_0 will be of decreasing type rel f^a.*

The proof of Theorem 6.3 is at most trivially different from that of Theorem 6.1. Theorem 6.4 follows from Theorem 6.3 essentially as Theorem 6.2 follows from Theorem 6.1.

THEOREM 6.4. *If $\alpha < f(P_0) < \gamma$ and if the critical point P_0 is of decreasing type rel $f_{\alpha a}$, P_0 will be of decreasing type rel $f_{a\gamma}$.*

7. Complimentary Regions and Boundary Manifolds.—Throughout this section the following notation will be used. The function f defined over $\sum_n$, $n > 1$, shall be non-degenerate with distinct critical values. The constants $\alpha < \beta < \gamma$, $\min f < \beta < \max f$, shall equal no critical value of f. Otherwise α, β, γ shall be fixed but arbitrary. The parameter a shall increase from α to β. Let M_i be the number of critical points of f of index i with values on the interval (α, β). Of these M_i points let P_i and Q_i be, respectively, the numbers of points of increasing and decreasing type

rel f_{aa}, Q'_{n-i} and P'_{n-i} these numbers rel $f_{a\gamma}$, and let p_i, q_i, r_i be, respectively, the numbers of these M_i points of increasing, decreasing and neutral types, rel f^a. Observe that

$$f_{aa} \cup f_{a\gamma} = f_{\alpha\gamma}, \qquad f_{aa} \cap f_{a\gamma} = f^a. \tag{7.1}$$

Set $d_i = q_i - Q_i$. These differences are non-negative by Theorem 6.3.

THEOREM 7.1. *The algebraic increments ΔR_i and ΔB_i in the numbers $R_i(f_{aa})$ and $R_i(f^a)$ respectively, as a increases from α to β satisfy the relations $(i = 0, 1, \ldots, n-1)$.*

$$\Delta(R_i + R_{n-i-1} - B_i) = d_i + d_{i+1} + d_{n-i-1} + d_{n-i} + r_i + r_{n-i-1} \geqq 0. \tag{7.2}$$

In accordance with section 4

$$M_i = P_i + Q_i, \qquad \Delta R_i = P_i - Q_{i+1} \qquad (i = 0, \ldots, n), \tag{7.3}$$

where $Q_0 = Q_{n+1} = P_n = 0$. Using the results of Theorem 5.1

$$\left.\begin{aligned} M_i &= p_i + q_i + r_i && (i = 0, 1, \ldots, n) \\ \Delta B_i &= p_i - q_{i+1} + p_{n-i-1} - q_{n-i} && (i = 0, 1, \ldots, n-1), \end{aligned}\right\} \tag{7.4}$$

where $p_n = q_0 = 0$ and $r_i = 0$ for $2i \neq n$. From (7.3) and (7.4), respectively

$$M_i + M_{n-i-1} = \Delta(R_i + R_{n-i-1}) + Q_i + Q_{i+1} + Q_{n-i-1} + Q_{n-i}$$
$$M_i + M_{n-i-1} = \Delta B_i + q_i + q_{i+1} + q_{n-i-1} + q_{n-i} + r_i + r_{n-i-1}$$

and (7.2) follows on equating these two representations of $M_i + M_{n-i-1}$.

COROLLARY 7.1. *If C_n is an arbitrary regular subregion of $\sum_n$*

$$R_i(C_n) + R_{n-i-1}(C_n) - R_i(BC_n) \geqq 0 \qquad (i = 0, 1, \ldots, n-1). \tag{7.5}$$

Choose f (as is possible) so that $f_\beta = C_n$ and choose $\alpha < \min f$. The left member of (7.2) becomes the left member of (7.5).

COROLLARY 7.2. *Necessary and sufficient conditions that $\sum_n$ have the connectivities of the n-sphere are that (a) for any non-degenerate f with distinct critical values each critical point with f-value $<$ max f be of the same type (increasing or decreasing) rel f^a as rel f_a, or (b) for each regular subregion C_n of $\sum_n$ the left members of (7.5) are zero.*

Choose $\alpha < \min f$ and β greater than each critical value less than max f. Then f^β is a topological $(n-1)$-sphere. If $\sum_n$ has the connectivities of the n-sphere the left members of (7.2) vanish. Hence each d_j and r_j is zero. The necessity of (*a*) follows. Conversely if (*a*) holds the right members of (7.2) are zero and hence the left. Choosing α and β as previously one sees that f_β has the connectivities of an n-cell and hence $\sum_n$ those of an n-sphere. Corollary 7.2 (*b*) follows trivially from Corollary 7.2 (*a*).

THEOREM 7.2. *If ΔR_i and $\Delta R'_i$ are, respectively, the algebraic changes in $R_i(f_{a\alpha})$ and $R_i(f_{a\gamma})$ as a increases from α to β, then for $i = 0, 1, \ldots, n-1$,*

$$\Delta(R'_{n-i-1} - R_i) = (Q_i - P'_{n-i}) + (Q_{i+1} - P'_{n-i-1}) \leqq 0. \quad (7.6)$$

The parentheses in the right members of (7.6) are non-positive by Theorem 6.4. The equalities in (7.6) follow from the relations

$$P_i + Q_i = P'_{n-i} + Q'_{n-i} \quad (i = 0, 1, \ldots, n)$$
$$\Delta R_i = P_i - Q_{i+1}, \quad \Delta R_i' = Q'_{i+1} - P'_i \quad (i = 0, 1, \ldots, n-1)$$

COROLLARY 7.3. *If C_n is an arbitrary regular subregion of $\sum_n$ then for $i = 0, 1, \ldots, n-1$*

$$R_{n-i-1}(\text{Comp } C_n) - R_i(C_n) \leqq R_{n-i-1}(\textstyle\sum_n). \quad (7.7)$$

The left member of (7.6) has the form

$$R_{n-i-1}(f_{\beta\gamma}) - R_i(f_{\alpha\delta}) - R_{n-i-1}(f_{a\gamma}) + R_i(f^a). \quad (7.8)$$

To obtain (7.7) from (7.6) choose f so that $f_\beta = C_n$. Take $\alpha < \min f$, $\max f < \gamma$. Then each $R_i(f^a) = 0$ and the remaining terms in (7.8) become those of (7.7).

COROLLARY 7.4. *If $\sum_n^\circ$ is a regular subregion of $\sum_n$ and if C_n and C'_n are complimentary regular subregions of $\sum_n^\circ$ then for $i = 0, \ldots, n-1$*

$$R_{n-i-1}(C'_n) - R_i(C_n) \leqq R_{n-i-1}(\textstyle\sum_n^\circ) - R_i[B(C_n) - (C_n \cap C'_n)]. \quad (7.9)$$

This is an obvious consequence of (7.6), using (7.8), on choosing f and α, β, γ (as is possible) so that

$$f^\alpha = B(C_n) - (C_n \cap C'_n), \quad f^\beta = C_n \cap C'_n, \quad f_{\alpha\beta} = C_n, \quad f_{\beta\gamma} = C'_n. \quad (7.10)$$

With $\sum_n$ as in Corollary 7.2 (*a*), a critical point P_0 of f with $f(P_0) < \max f$ is of the same type (increasing or decreasing) rel f_a as rel f^a, and hence, if $f(P_0) > \min f$, of the same type (increasing or decreasing) rel F_a as rel F^a. A review of the proof of Theorems 6.2 and 6.4 then gives the following.

THEOREM 7.3. *If $\sum_n$ has the connectivities of the n-sphere a critical point of f of index k with $0 < k < n$ is of the same type (increasing or decreasing) rel $f_{a\alpha}$ as rel $f_{a\gamma}$, and (7.6), (7.7) and (7.9) become equalities for $i = 1 \ldots, n-1$.*

In particular if $\sum_n$ has the connectivities of the n-sphere the equality (7.7) gives a combinatorial extension (1927) of Alexander's result

$$R_{n-i-1}(\text{Comp } C_n) = R_i(C_n) + \delta^i_{n-1} \quad (i = 1, \ldots, n-1). \quad (7.11)$$

[1] Morse, M., "The Analysis and Analysis Situs of Regular n-Spreads in $(n+r)$-Space," PROC. NATL. ACAD., **13**, 813–817 (1927).

[2] Morse, M., "Relations between the Critical Points of a Real Function of n Independent Variables," *Trans. Am. Math. Soc.*, **27**, 345–396 (1925).

[3] Morse, M., and Van Schaack, G. B., "The Critical Point Theory under General Boundary Conditions," *Ibid.*, **35**, 545–571 (1934).

[4] Morse, M., "The Calculus of Variations in the Large," *Am. Math. Soc. Coll. Publ.*, New York, 1934.

[5] Seifert-Threlfall, *Lehrbuch der Topologie*, Teubner, Leipzig and Berlin, 1934.

CRITERIA FOR LINEAR EQUIVALENCE

By André Weil

University of Chicago

Communicated by Oscar Zariski, January 5, 1952

1. Equivalence criteria play an important part in the theory of algebraic surfaces; such criteria have been established by Severi[1] for surfaces without multiple points over the field of complex numbers. My work on abelian varieties[2] has enabled me to extend Severi's methods and results to arbitrary normal varieties in projective spaces. As publication of the complete proofs, which I have already written up, may be somewhat delayed, I shall give here a precise statement of the main results which have been so obtained.

Once for all, V will denote a variety of dimension n without multiple subvarieties of dimension $n - 1$, embedded in a projective space P of dimension N and not contained in any hyperplane. Let k_0 be the smallest field of definition for V; all the fields to be mentioned will be assumed to contain k_0. If H is a hyperplane in P, generic over k_0, then[3] the cycle $V.H$ is a variety of dimension $n - 1$ without multiple subvarieties of dimension $n - 2$. If L is a linear subvariety of P of dimension $N - n + 1$, generic over k_0, then $V.L$ is a curve C without multiple points which does not go through any multiple point of V; such a curve $C = V.L$ will be called a *general curve* on V; if, moreover, L is generic over a field k containing k_0, then the general curve C will be said to be generic over k.

2. Linear equivalence of divisors on V is defined as usual, and will be denoted by $\sim$. Let W be an abstract variety; let Z be a divisor on $V \times W$; if M is a simple point on W such that the cycle $Z.(V \times M)$ is defined, we write $Z.(V \times M) = Z(M) \times M$. Any divisor of the form $Z(M) - Z(N)$, where Z is as just stated and M, N are such that $Z(M)$, $Z(N)$ are defined, is said to be *algebraically equivalent to 0*; one of my lemmas states that such divisors form a group and that every such divisor can be written as $Z(M) - Z(N)$ by means of a divisor Z on the product $V \times \Gamma$ of V and of a curve Γ.

Criteria for linear equivalence can be of two kinds; those of the first kind state that a divisor X is linearly equivalent to 0 provided its trace on a subvariety of V which depends upon X is linearly equivalent to 0 on that subvariety; those of the second kind state that a divisor X is linearly equivalent to 0 provided it is algebraically equivalent to 0 and its trace on a fixed subvariety of V is equivalent to 0 on that subvariety.

3. The criteria of the first kind are as follows:

(A) Let X be a divisor on V, rational over a field k; put $W = V.H$, where H is a hyperplane in P, generic over k. Then, if $n \geqq 3$, $X.W \sim 0$ on W implies $X \sim 0$ on V; the same is true for $n = 2$ if one assumes at the same time that there is an integer $m \neq 0$ such that $mX \sim 0$ on V.

For $n = 2$, $X.W \sim 0$ need not imply $X \sim 0$ because of the exceptional surfaces occurring in the theorem of Kronecker-Castelnuovo; what is more, there are serious difficulties in the way of extending this theorem to arbitrary fields.[4] Nevertheless, we have the following general result:

(B) There is a finite set of divisors D_i on V, all algebraic over k_0, with the following properties: (i) no linear combination of the D_i, other than 0, is algebraically equivalent to 0; (ii) let X be a divisor on V, rational over a field k; let C be a general curve on V, generic over k; then $X.C \sim 0$ on C if and only if X is linearly equivalent on V to some linear combination of the D_i.

4. Let f be a mapping of V into an abelian variety A; if X is a divisor on V, and C a general curve on V such that $X.C$ is defined, I define $f(X)$ by putting $f(X) = S[f(X.C)]$, where S is as defined, *loc. cit.*,[2] pp. 28–29; this is independent of the choice of C; if k is a field of definition for V, A and f, and if X is rational over k, $f(X)$ is rational over k. If $Z(M)$ is as defined above in paragraph 2, there is a mapping g of W into A such that $g(M) = f(Z(M))$ whenever $Z(M)$ is defined.

The criteria of the second kind can now be formulated:

(C) Let X be algebraically equivalent to 0 on V. Then, if $f(X) = 0$ for every mapping f of V into an abelian variety, there is an integer $m \neq 0$ such that $mX \sim 0$ on V.

(D) Let W be an abstract variety; let Z be a divisor on $V \times W$. Assume that, for every mapping f of V into an abelian variety, the mapping g of W into the same variety defined by $g(M) = f(Z(M))$ is constant. Then all the divisors of the form $Z(M)$ are linearly equivalent to one another on V.

(E) Let C be a general curve on V. Then, if a divisor X is algebraically equivalent to 0 on V and $X.C$ is defined and ~ 0 on C, X is linearly equivalent to 0 on V.

Only one consequence of the above results will be mentioned here: If k is any field of definition for V, and a divisor X_0 is rational over k, then every divisor X satisfying a relation $mX \sim X_0$, where m is an integer other than 0, is linearly equivalent to some divisor X_1 which is algebraic over k.

On the other hand, it will be clear to everyone who is familiar with this subject that the above results, and particularly (*E*), open the way to a theory of the so-called Picard varieties of arbitrary varieties.[5]

[1] Severi, F., *Ann. di Mat.* (III), **12**, 55–79 (1905); *Rend. Pal.*, **21**, 257–282 (1906); *Rend. Acc. Lincei* (V), **30**, 328–332 (1° sem. 1921). Cf. also O. Zariski, *Algebraic Surfaces*, Chelsea, New York, 1948, pp. 88–89 and pp. 126–128.

[2] Weil, A., *Variétés abéliennes et courbes algébriques*, Act. Sc. et Ind. no. 1064, Hermann et Cie, Paris, 1948.

[3] Matsusaka, T., *Kyoto Math. Mem.*, **26**, 51–62 (1950).

[4] I am, however, informed by P. Samuel that he has been able to avoid all such difficulties by the use of a suitable birational transformation.

[5] Weil, A., *Colloque d'Algèbre et Théorie des Nombres*, Centre Nat. de la Rech. Scient., Paris, 1950, pp. 125–127. I have been informed by T. Matsusaka, and also by A. Néron and P. Samuel, that they have independently developed theories of the Picard varieties. By means of the device referred to in footnote 4, Néron and Samuel have been able to derive the basic results on the existence of the Picard variety from the most elementary form of the criterion of the first kind.

RAPID EFFECTS UPON THE RENAL CIRCULATION PRODUCED BY NEPHROTOXIC GLOBULIN ADMINISTRATION IN THE RAT

BY RICHARD W. LIPPMAN, HELEN U. MARTI AND E. ELMO JACOBS

INSTITUTE FOR MEDICAL RESEARCH, CEDARS OF LEBANON HOSPITAL, LOS ANGELES, CALIFORNIA*

Communicated by Linus Pauling, January 28, 1952

Although rabbit anti-rat-kidney serum has been shown to contain antibodies that localize in the glomeruli,[1] renal tubule cells also are damaged by the administration of this serum or of rabbit anti-rat-kidney gamma globulin (nephrotoxic globulin, NTG). Indeed, the damaging effect of NTG on renal tubule cells has been shown in tissue culture explants, where indirect circulatory effects are excluded.[2] It has been suggested that the tubular damage produced *in vivo* by administration of NTG might result from either a direct effect upon the tubule cells or circulatory disturbances subsequent to NTG administration, or from a combination of these two factors.[3]

Various methods have been used to study circulatory changes in the kidney. The injection of foreign materials, such as neoprene or India ink, may produce distortions of the vascular pattern, as a result of mechanical or pharmacologic effects. Such a quasi-physiologic method as the measurement of renal blood flow by means of para-aminohippurate or diodrast clearances cannot be used after NTG administration, since the

animals are anuric for at least 2 hours, a period of critical importance. The fluorescent dye vasoflavine presents technical difficulties which we have not mastered with entire satisfaction.

For these reasons, we have now devised a very simple method of examining the renal circulation. The rat was placed under light ether anesthesia, the abdomen was opened widely, the abdominal viscera were retracted gently with pads of cotton dipped in warm 0.85% sodium chloride solution, and a kidney was exposed to direct observation. The solution under investigation was then administered by intravenous injection, in a foot vein, and the kidney was observed for changes in the surface appearance. At any desired time the renal pedicle was rapidly and tightly occluded by application of a clamp, thus trapping blood in the renal vascular bed, since the artery and vein were simultaneously occluded. The kidney, with clamp on pedicle, was rapidly removed *en bloc* and a ligature was placed around the pedicle before removing the clamp. The entire kidney was then dropped into 10% neutral formalin in 0.85% sodium chloride solution. After 24 hours the kidney was sliced in half, and fixation was continued for another 24-hour period. Paraffin sections were made at a thickness of 15 μ, and were stained by the benzidine method of Ralph.[4]

Control globulin (GC) and NTG were prepared by the methods previously described in detail.[3, 5] In the control animals 1.0 ml. of GC (17.0 mg. total protein) was administered by intravenous injection. In the other animals 1.0 ml. of NTG (23.0 mg. total protein) was administered. For this experiment 39 rats of the Slonaker-Addis strain[6] were used.

In the 8 male, 150-g. animals given GC no observable change in the surface appearance of the kidney could be seen during 2 minutes after the globulin injection. The kidney was clamped and excised, after the 2-minute period of observation, and sections were made for histologic examination. It was noted that the glomeruli contained erythrocytes in the capillary channels and, except for occasional rouleaux, their number was such that they could be identified as discrete bodies. Small numbers of erythrocytes could be seen in the peritubular capillaries and venules. Occasional dilated venous channels could be seen in the cortical area, and, adjacent to these, the peritubular capillaries appeared to contain more blood than vessels in other areas. Blood could also be seen in the major vessels and in the vasa recta, extending toward the pelvic papillae.

After the administration of NTG a dramatic succession of events was observed on the surface of the kidney, *in situ*. Within 15 to 45 seconds after the injection was given, small livid areas, similar to petechiae, were seen on the exposed renal surface. These were so deeply red in color that, at times, the intervening renal tissue appeared to be blanched. This appearance was only by comparison with the deeply colored areas, and no true blanching occurred. The "petechiae" rapidly increased in number

until there was a gradual fusion of these livid areas, and the entire kidney became a deep, meaty red, but this color was not cyanotic. The complete change occurred within 2 minutes in most instances, with only an occasional animal that required 3 minutes. This experiment was performed on 19 male, 150-g. animals and, of these, in 7 the kidney was clamped and excised at intermediate stages, while in 12 the kidney was removed after 2 minutes or slightly more, when the gross surface change had become maximal. When sections were examined, it was found that the glomeruli contained many more erythrocytes than were contained in the glomeruli of animals which had received GC. Many more rouleaux were present, and the erythrocytes were at times so closely packed that it was impossible to identify them discretely. The number of dilated venous channels in the cortical area was increased, and there was an increase in the amount of blood contained in the peritubular capillaries and venules. In addition the vasa recta contained more blood than in the control specimens.

For comparison, 6 specimens were obtained in which the renal vein was isolated and ligated for 2 minutes prior to application of the pedicle clamp and removal of the kidney. In addition, in 6 other animals the renal artery was ligated for 2 minutes prior to application of the clamp and removal of the kidney. In specimens in which the renal artery had been ligated the entire kidney appeared to be almost devoid of blood, with a few isolated erythrocytes in each glomerulus and with very little blood in the peritubular capillaries and other vessels. In specimens obtained from animals in which the renal vein had been ligated there was massive congestion of the venous channels, including the peritubular vessels, but there was no significant change in the glomerular blood content.

While the technique described here does not exclude the possibility of intrarenal shifts in blood distribution during the period between application of the clamp and penetration of the fixative, the likelihood of a distorted picture is very much less than in most other methods that have been used to study the renal circulation. It is of interest to note that, in any given kidney section, there was some variation in the blood content between different radial segments. The experimental differences, however, far exceeded the variability observed in individual sections.

The most remarkable observation in this experiment was the speed with which NTG administration affects the renal circulation. Since the mixing time in the rat is approximately 2 minutes,[7] and the vascular changes observed in the kidney begin after only 15 seconds, it would seem probable that such changes occur almost instantaneously when NTG arrives at the kidney. In previous work the speed with which such major changes occur has not been appreciated, and observations have usually been made at much longer intervals, ranging from 30 minutes to 1 week after the administration of NTG.

Obstruction to the venous outflow from the kidney, as demonstrated here, fills the peritubular vessels but does not appreciably alter the glomerular blood content. After NTG administration, the glomeruli, as well as the peritubular and venous channels, contain an increased amount of blood. Since efferent arteriolar constriction would increase the glomerular blood content but probably would diminish the amount of blood in venous channels, it would seem reasonable to suppose that the events observed are explained by an increased amount of blood reaching the kidney from the arterial side, rather than by obstruction to the venous outflow at any point. This idea is confirmed by the color change which, when grossly observed, does not have the appearance of cyanosis, which one would expect with venous congestion alone.

Exploratory experiments performed with India ink injection and vasoflavine, in spite of the technical difficulties, have given results that confirm the information obtained by the much simpler method described here.

In the light of these experiments, it seems clear that the administration of NTG does not result in tubular ischemia. Additional preliminary experiments have shown that the renal hyperemia observed here lasts for some time, certainly more than 15 minutes, although possibly less than 1 hour. It would seem more reasonable to believe that the tubular damage after NTG administration is the consequence of a direct, noncirculatory effect upon the tubular cells.

* This work was aided by a grant from the National Heart Institute, of the National Institutes of Health (H 132 C3), and by a grant from Ciba Pharmaceutical Products, Inc., Summit, N. J. The authors are grateful for the technical assistance of Ruth Taniguchi. Dr. Lippman is a Fellow of the John Simon Guggenheim Memorial Foundation.

[1] Pressman, D., Hill, R. F., and Foote, R. F., *Science*, **109**, 65 (1949).

[2] Lippman, R. W., Cameron, G., and Campbell, D. H., PROC. NATL. ACAD. SCI., **36**, 576 (1950).

[3] Lippman, R. W., Marti, H. U., and Campbell, D. H., *Arch. Path.* **53**, 1 (1952).

[4] Ralph, P. H., *Stain Technol.*, **16**, 105 (1941).

[5] We are deeply grateful to Prof. Dan H. Campbell and Jay Banovitz, Gates and Crellin Laboratories of Chemistry, California Institute of Technology, for preparation of the NTG and GC used in these experiments.

[6] Addis, T., and Gray, H., *Growth*, **14**, 49 (1950).

[7] Lippman, R. W., *Proc. Soc. Exp. Biol. Med.*, **66**, 188 (1947).

SHAPE AND MOVEMENT OF MESENCHYME CELLS AS FUNCTIONS OF THE PHYSICAL STRUCTURE OF THE MEDIUM. CONTRIBUTIONS TO A QUANTITATIVE MORPHOLOGY*

BY PAUL WEISS AND BEATRICE GARBER

DEPARTMENT OF ZOOLOGY, UNIVERSITY OF CHICAGO

Communicated January 29, 1952†

Introduction.—Historically, the description and classification of cells has proceeded from visual criteria. Thus shape became the predominant feature, and even cytological stains, which are actually cytochemical tests by nature, were used primarily for their value in making cellular detail visible or more distinct under the microscope. It is the task of contemporary biology to replace such static and purely formal descriptions of living systems by dynamic concepts which will define these systems objectively by reference to their inherent properties expressed in measurable standard units, rather than in terms of subjective impressions. To quote from an earlier paper (Weiss, 1950), "We must form the habit, for instance, of viewing the shape of a given cell, tissue, or organ not as a static feature, but as a phase, transitory or terminal, in a continuous chain of transformations; as a cross-section through a stream of processes proceeding in time. Every morphological criterion is but an expression of antecedent processes that have contributed to its formation. Distance, in this version, becomes an index of past attractions, separations, intercalations, etc.; shape, other than spherical, the result of distortions due to elongation, secretions, shifts, contractions, expansions, differential growth, and the like; . . ."

In this view, morphological appearance becomes but an incidental sign of prior morphogenetic transformations, and cell shape can be understood only in terms of the chains of interactions through which it has come about. Form thus loses its connotation of absoluteness and assumes the role of a mere sample taken during the course or at the end of a continuous formative process. To translate organic form into such objective and dynamic terms —"morphostatics" into" morphodynamics"— will take quite some time and effort. In the meantime, we shall have to continue, for convenience, to use such picturesque designations as fusiform, stellate, arborized, crenated, flame cells, goblet cells, etc. But our ultimate goal must remain to express them (*a*) in quantitative terms, and (*b*) in terms of their genesis. The classical case of a program of this sort has been stated by d'Arcy Thompson (1942) in his book on *Growth and Form*. The present paper offers a small contribution to this program. It will mainly sketch the theoretical premises of our work and some illustrative experimental results, while the detailed report of the experiments will follow in a later paper by Garber.

Cell shapes can be classified according to causes in two distinct groups depending on whether they are due to (1) passive deformation from the outside, or (2) active transformation by cellular activities. Examples of (1) are the configuration of cells closely packed into a confined space (e.g., blastomeres; epithelial cells, F. T. Lewis, 1933; plant meristem, Matzke, 1947), following general geometrical rules and biologically of relatively minor interest. Class (2) can be subdivided into two types, namely, (*a*) deformations within a cell caused by its own internal products (e.g., local accumulation of secretion granules; fibrous protrusions, such as flagellae, sperm tails, etc.; crystalline endoskeletons, W. J. Schmidt, 1924); and (*b*) transformations of the whole cell body into elongate, discoid,

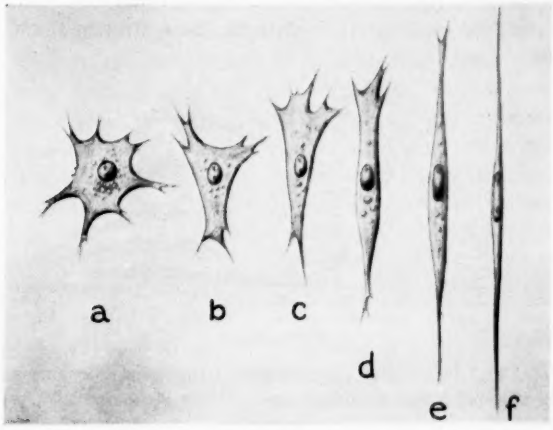

FIGURE 1

Series of cell forms observed in fibroblast cultures *in vitro* semidiagrammatic).

ramified, etc., shapes (e.g., mesenchyme cells, neurons, chromatophores, glia cells). It is with this last group that we shall deal in the following.

Figure 1 shows a graded series of cell samples commonly encountered in mesenchyme and connective tissues, both in the body and in tissue cultures. Observation has shown that these are widely interconvertible, that is, the same cell can assume one or another of these guises depending on circumstances. We shall subject this series to morphogenetic analysis in the outlined sense. This involves three steps. First, we must express the diverse shapes in appropriate parameters that will permit us to range them along a single quantitative scale. Next, we must try to correlate the observed morphological variations with some systematic variations either

within the cells or in the environment. Lastly, having succeeded in this, we must determine the physical basis, the "mechanism," underlying this correlation; for it is one thing to ascertain, let us say, that a cell moves the faster, the thinner the medium (e.g., Willmer,1933), and another thing to explain just why.

The direction in which to look for such an explanation was clearly pointed by earlier observations (Weiss,1929), in which the two extreme forms of the above series (Fig. 1) were found to be characteristically associated with fluid-rich and highly condensed plasma media, respectively. Subsequent work (Weiss, 1934, 1944a, 1945; Weiss and Wang, 1944) confirmed the dependence of cell shape and cell movement on the physical structure of the medium. This is schematically summarized in the following diagram (reproduced from Weiss, 1949), which represents crudely the fibrin net of a clot, the lower part of which has been stretched, and which con-

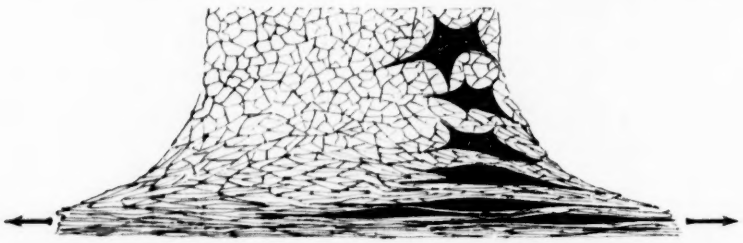

FIGURE 2

Effect of stretch (arrows) on the structure of the fibrin reticulum of a plasma clot and on the shape of cells contained therein. (From Weiss, 1949).

tains mesenchyme cells in its different regions (Fig. 2). Where tension has aligned and condensed the fibrin (lower part), cells are bipolar and elongate in the direction of the fibrin strands, while in the relaxed portions (top), cells are multipolar and plump, with intergrades in the intermediate zones. These cell forms evidently correspond closely to the series pictured in figure 1. Yet, since the whole variety can also be observed, often side by side, in unstretched clots, one had to search for a more comprehensive explanation. The clue in this search was that the variety of cell forms was clearly correlated with some variable property of the medium, evidently one that varied systematically with stretch, but at random in ordinary clots.

Cell Parameters.—For a quantitative formulation of cell shape, we choose its deviations from the simplest form, the spheroid (which most cells resume prior to mitosis and after death). Ignoring the third dimension, all forms of figure 1 can be derived from a sphere by extending certain

radii, varying in number and distribution: in *a*, there are 7 or 8 such major projections in relatively uniform distribution along the surface (multipolar cell); in *b*, there are 3 major ones, about equidistant (tripolar cell); in *c*, 3 major ones, of which two are almost antipodal to the third (pseudobipolar cell); and in *d*, *e*, *f*, 2 major ones (bipolar cell). The number of major processes thus furnishes us with one quantitative scale. As the number of processes decreases, bipolarity becomes more marked, and it becomes possible then to distinguish between length and width of the cell. We shall use the ratio of length over width as index of cell elongation. The shape of the similarly elongated nucleus can likewise be expressed by the ratio of longest over shortest nuclear diameter. These three variables will be designated in the following as N for number of major processes; E_c for elongation of the cell body and E_n for elongation of the nucleus.

Factors of Deformation.—Deviation from sperical shape involves shifts of the cell content, with some parts extending, others receding. For instance, when a cell attaches itself to a flat surface, its border tends to expand in practically all directions at the same time, thus flattening the mass to a disk at the expense of height (cf. Weiss, 1944a). Forms such as those in figure 1, on the other hand, result from the extension of the cell surface at relatively fewer points, where tongues of cytoplasm advance centrifugally, while the intermediate areas sag, assuming catenary contours. The active centrifugal movement in the elongating radii produces radial stresses, while the sagging parts of the surface come under tangential stress. These stress patterns can sometimes be made visible by impregnating the correspondingly aligned fibrous units in the cytoplasm with silver (e.g., Fig. 12 in Weiss, 1944a) or by observation under the electron microscope (Porter, Claude and Fullam, 1945) and they are valuable clues to the mechanism of deformation; they are schematically indicated in figure 3.

The protrusion of any part of the surface is opposed by the leveling forces of surface tension and elasticity. Therefore, in order for any protrusion to last more than just momentarily, the force and rate of its thrust must be high enough to overcome these opposing forces. Evidently, the average number of projections to be expected from any given cell will depend on the incidence of such break-throughs.

In surveying different cell types, it is noted that those listed above as Class 2*a* can often project long arms into structureless (liquid) media by virtue of endocellular fibrous or crystalline axial filaments. Unlike these, cells of Class 2*b*, with which we are here concerned, are devoid of solid endoskeletons and seem to be unable to form long processes without the aid of an outside support—generally speaking, of an interface—to which they can cling (cf. Weiss, 1945). Surface protoplasm is drawn out along such

interfaces by adhesive forces which vary with the gross physical properties and the subtler molecular relations along the contact surfaces and which may be lumped here for brevity under the common term of "interfacial tensions." Along a flat interface, marginal protoplasm will thus be drawn out as a thin veil with a broad front (W. H. Lewis, 1950; Weiss, 1944a), while along a cylindrical interface, e.g., the surface of a fiber, it will be drawn out into a pointed filament. Without going into detail, it should be mentioned that the forces needed for the formation of such a cell process seem to be provided partly by the interfacial pull and partly by intracellular metabolic energy, neither of them sufficient to do the work alone (see below). Perhaps the interface, by orienting the movement of the protoplasmic surface, just aligns crucial molecular chains, and metabolic energy then links them by polymerization into a framework consistent enough to withstand retractive forces.

Be this as it may, the morphogenetic significance of the interface as such is well substantiated. Hence, any fiber of the extraneous medium that makes contact with, or pierces, the cell surface, will call forth a little pointed projection, vaguely comparable to the meniscus rising on a wire partly immersed in a wetting liquid.

Now, once established, such a protrusion acts as a viscous flow, to which hydrodynamic considerations may be applied. That is, it exerts a draining suction on its surroundings in proportion to its momentum. In consequence, two or more projections extending concurrently from the same surface cannot proceed independently, but are in competitive interaction with one another, as well as with the cohesive and elastic forces of the system. They will contend for the influx of protoplasm among themselves according to their relative momentums, with the larger and stronger ones tending to drain (hence resorb) the smaller and weaker ones. The application of this principle of tug-of-war among cell processes to the mechanics of nerve fiber growth and branching has been presented previously (e.g., Weiss, 1941, 1944b). A general consideration of systems of this order leads to the following rules:

(1) An advancing process tends to prevent the emergence of another process within a certain radius around its base by competitive drainage. (2) This inhibitory effect declines with increasing distance from the base of the advancing column. (3) The effect increases with increasing mass and velocity of the advance. (4) Other processes already in existence within the drainage area can proceed if their mass and velocity are large enough to entertain a counter-drain. Otherwise they will be resorbed.

To these four dynamic rules must be added a fifth one based on a micromorphological consideration. If we assume that the formation of a cell process along an extraneous fiber involves the aligning of fibrous units in the cell in the direction of the advance, the probability of the emergence

of a new process at a given contact point along the cell surface will vary inversely with the difficulty with which molecular realignment at that point can be brought about. The difficulty will be least where orientation is already predominantly perpendicular to the border (e.g., at the advancing front of a stream) while it will reach a maximum where the orientation is parallel to the border, particularly if the elements have become linked into larger surface-parallel chains. Rule (5) therefore states that a contact stimulus of given magnitude will be more likely to produce a local "microlesion"

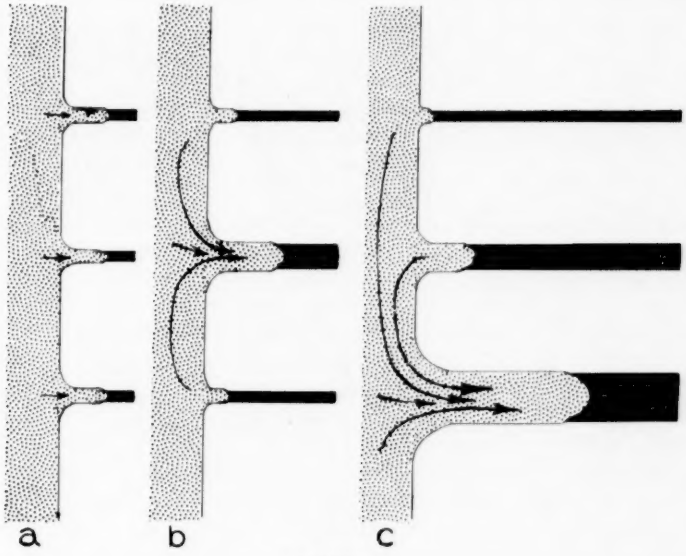

FIGURE 3

Reaction of the cell surface (stippled) to contact with extraneous fibers of different sizes.

or breach of the cell surface if it is applied near the tip, than if it is near the flank, of a moving cell part. As reported in an earlier paper (Weiss, 1944a), observations on phagocytosis point to the same conclusion, as granules are ingested only along those parts of the cell border that seem to possess a radial microstructure, but not across zones with tangential orientation (for supporting evidence, see also Hancox, 1949).

Relation between cell processes and texture of medium.—The five cited rules permit us to predict the conditions making for greater or smaller average numbers of processes in given cell types in a medium containing net-

works of fibers such as the fibrin reticulum of a blood plasma clot. Figure 3 represents schematically small samples of three cell surfaces (cytoplasm stippled), each in contact with three equidistant fibrin threads (black) of identical (*a*) or different (*b*, *c*) size. In line with what was said above, we conclude that at each contact point a protoplasmic process will be sucked forth by interfacial tension, the draining force being proportional to the size of the fiber. Larger fibers thus have a competitive advantage over smaller ones. In figure 3*a*, all three fibers being of equal size, equilibrium is attained with the establishment of three processes. In figure 3*b*, the larger middle fiber weakens the processes along the two

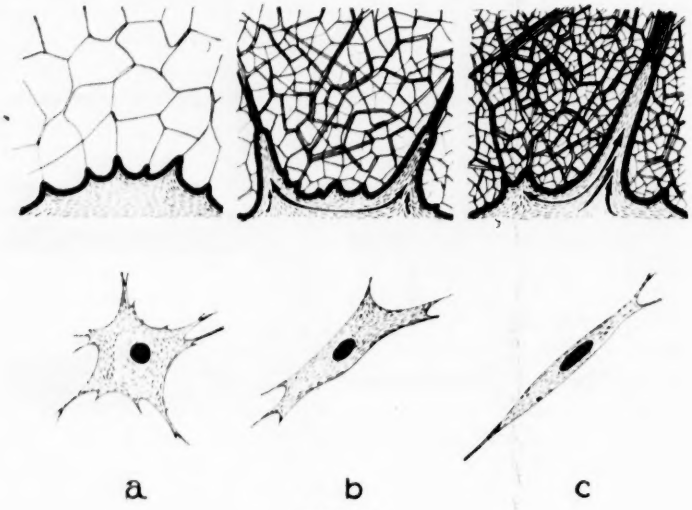

FIGURE 4

Morphology of mesenchyme cells in three media of (from left to right) increasing plasma concentrations.

outer ones. In figure 3*c*, the largest outflow along the bottom fiber will drain both the middle and the top processes. We may conclude then that for a given density of the fibrous fabric, the number of cell processes that can be sustained per unit surface will be the larger, the more uniform the fiber diameters are, and the smaller, the greater the diversity of fiber sizes. Furthermore, according to rule (5), the stronger the oriented flow (the larger the fiber), the smoother will be the edges. In consequence, the number of processes should vary inversely with the proportion of large fibers in the medium.

To test this assumption, we studied cell shapes in plasma clots prepared in a graded series of fibrin textures. The reactions to be expected are

illustrated in figure 4, showing in the upper row, a small segment of a cell border exposed to three clots of mounting fibrin concentrations, and in the lower row, the shapes displayed by the respective cells (for simplicity, cytoplasm and fibrin are not drawn to the same scale). As fibrin concentration increases from *a* to *c*, more and more elementary filaments combine into larger fibers, resulting in an increase of both the range of the fiber spectrum and the proportion of large fibers. These three diagrams, corresponding essentially to the three phases of figure 3, summarize the conclusion drawn above, that the higher the incidence of large fibers in the medium, the greater is the probability that the cell surface will be vigorously drawn out along one of them, thus automatically suppressing the formation of other sprouts for some distance. The diagrams thus lead us to expect an inverse statistical relation between coarseness of medium and average number of major cell processes.

Stated succinctly, the principle involved is as follows. Small fibers evoke weak protrusions with a small draining radius, hence, able to coexist without mutual interference even at close range. Larger fibers, evoking a larger outflow, establish a wider drainage field. Multiple extensions within such a common field, being in hydrodynamic and elastic competition, could only persist if all of them were of exactly identical strength, which would presuppose uniformity in the size of the guide fibers. The greater the range of guide fiber sizes, the less will be the chance for such an equilibration to occur, hence, the greater will be the probability that a single process will assume dominance. It is for this reason that we must expect the number of cell processes to be a function not only of maximum fiber size but also of the range of fiber sizes in the medium.

Experimental Tests.—Media of controlled texture were obtained by two different methods. In one set of experiments, culture media were made up of mixtures of fowl plasma and chick embryo extract in varying ratios, ranging from 10 to 90% plasma concentration. With increasing concentrations, the tendency of the fibrin filaments to aggregate to larger bundles rises steadily, thus providing us with textures of increasing coarseness. Although this series gives us the desired physical gradations, it also introduces chemical differentials due to the varying ratios of plasma and embryonic substances. Therefore, in a second set of experiments plasma and extract were mixed in equal amounts (40% plasma concentration), but coagulated at different pH over a range of from 5.6 to 8.0. It is known and has been crucially confirmed by the electron micrographs of Hawn and Porter (1947) that the lower the pH, the coarser become the bundles of fibrin. This then provides us with a means of grading physical texture while keeping chemical composition constant. Since both series gave identical results regarding cell shape, physical texture is proved to be the relevant variable.

The cultures were standard heart fibrocyte cultures having been passed through five transfers to obtain cell strains of relatively uniform character. Clots coagulated at unphysiological pH values were quickly buffered back to normal by the growing tissue. Because surface and interior of a clot are vastly different in texture, only cells in the depth of the clot were recorded, excluding any that had spread along the coverslip or the air interface. Cells were measured in camera lucida drawings of either the living cultures or the fixed preparations. Samples were taken at random. The ranking of a given process as major (i.e., affecting cell shape significantly) or minor, is, of course, somewhat arbitrary, but since large numbers of cells were used and the same standards were applied in the rating throughout, the results of the different series were comparable and consistent. In bipolar cells, length was measured from tip to tip; width, across the cell body at the level of the nucleus. The following results are based on measurements of a total of 4759 cells.

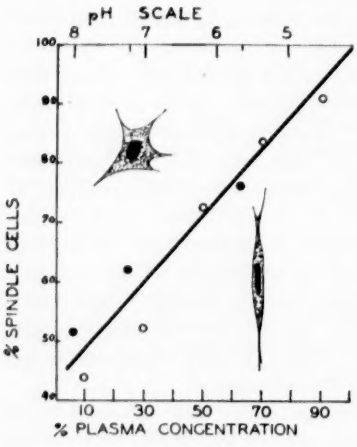

FIGURE 5

Proportion of spindle cells in plasma clots of different textures.

General Cell Shape.—Figure 5 expresses the most general relation between cell morphology and structure of the medium, namely, an increasing proportion of elongate (spindle) cells (types *d*, *e*, *f*, of Fig. 1) in the cell population with increasing coarseness of texture. The open circles give the percentages of spindle cells for the five plasma concentrations tested (lower abscissa), fitting a straight line which extrapolates to 100% spindle cells for a theoretical plasma concentration of 100%. The proportion of stellate cells (types *a*, *b*, *c*, of Fig. 1) decreases correspondingly. In media that had been coagulated at different pH (upper abscissa), the proportion of spindle cells likewise rises steadily with increasing coarseness (decreasing pH), and the resulting values (full circles) also roughly fit a linear function. The two curves representing the concentration and the pH functions have been brought to coincidence in the graph by an appropriate choice of abscissa units. The fact that as good a coincidence as is evident could be obtained proves that the relevant variable in determining cell morphology is really the texture of the medium without regard as to whether this is accompanied by variations in chemical concentration (lower

scale) or not (upper scale). The validity of this conclusion is further confirmed by the fact reported below that if other cell parameters are plotted in the same coordinate systems as those used in figure 5, they also fit a common curve. According to these data, therefore, as far as effects on cell morphology are concerned, the texture of a 10% plasma clot at normal pH is equivalent to that of a 40% plasma clot at pH 8, and an 80% clot at normal pH corresponds to a 40% clot at pH 5.

Number of Cell Processes.—Figure 6 presents a more detailed breakdown according to shapes of cell populations (2417 cells) in clots of graded plasma concentrations. On each ordinate are added the percentages of cells with 2, 3, 4 and 5 major processes. It can readily be seen that as plasma concentration rises, the average number of cell processes (N) declines steadily, again essentially in linear order and with a remarkable degree of regularity. The data thus bear out fully the prediction derived above from theoretical considerations.

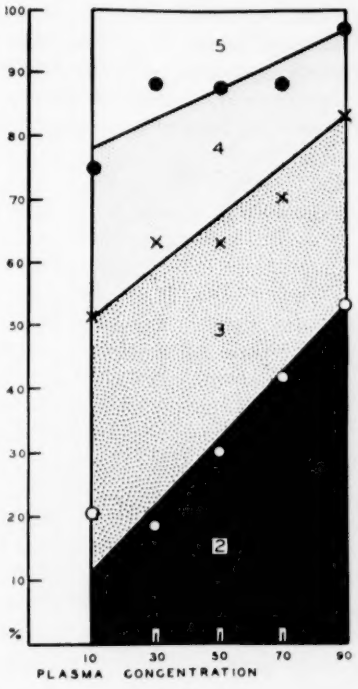

FIGURE 6

Proportions of cells with 2 (black area; white circles), 3 (dense stippling; crosses), 4 (light stippling; dark circles), and 5 (white) major processes in clots of different textures.

Elongation.—Parallel with the reduction of major cell processes goes a progressive elongation of the cell, leading from multipolarity through pseudobipolarity to strict bipolarity and then continuing, within the bipolar range, to increasing slenderness (Fig. 1). The quantitative expression of this transition is a steady increase in the index of elongation (E_c), i.e., the ratio of length over width, of the cell body. The nuclear proportions (E_n) conform to those of the cell body. Both series of values are plotted in figure 7 in relation to clot texture; open circles referring to plasma concentration (lower abscissa), full circles to pH at coagulation (upper abscissa). Clearly, the values of E_c and those of E_n define each a single continuous function of non-linear character. It will be noted that the numerical relation between the concentration and pH scales is the

same as in figure 5. The fact that with this setting, the values of both experimental series fit the same curve, justifies the assumption that lower and upper abscissae are truly interconvertible in point of morphological effects of clot texture.

These are empirical curves. In view of the complex factors involved, it would be idle to try, at this early date, to interpret their detailed shape in physical terms. Yet, their general character can be derived from our theoretical concept. A bipolar spindle cell may be regarded essentially as two processes drawn out along a common fiber in opposite directions, with one tip leading, the other trailing, and the rest of the cell body stretched out between them. According to our assumption, the pulling force is a function of interfacial tensions along the leading edge, which is the front line of contact between protoplasm and fiber. This force increases in some direct proportion to the circumference, hence the radius, of the fiber. Elongation results from the lag of the rear process in following the advance. This lag is due, among other things, to the adhesion of the rear process to the fiber, such adhesion being a function of the size of the contact area, hence, proportional to the square of the radius of the fiber. Therefore, since the advance increases linearly with the radius, while the drag increases as the square of the radius, elongation resulting from the disparity between rate of advance at the tip and rate of detachment at the rear, will be the greater, the larger the diameter of the fiber. Consequently, media with higher plasma concentrations, or coagulated at lower pH values, hence containing larger proportions of heavy fibers, must also be expected to yield cells with higher average E_c values, which is precisely what the curves of figure 7 demonstrate.

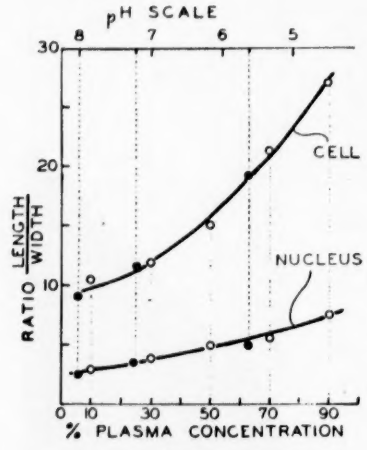

FIGURE 7

Ratios of length over width for cell bodies and nuclei in clots of different textures.

An unexpected feature revealed by these curves is the remarkable similitude between cellular and nuclear proportions. The ratio $E_c:E_n$ has an almost constant value for all observed cell shapes, averaging 3.4, with a range of from 3.1 to 3.8. This expresses a rather strict conformance of nuclear shape to the proportions of the cell body, but the mechanism by which this is brought about is not at all clear. It is apparent that

the elongation of the nucleus is directly and quantitatively related to the polarized structure and movement of the surrounding cytoplasm, but in view of our great ignorance regarding the factors determining nuclear shape in general, it remains uncertain whether this is a matter of passive deformation or of an active nuclear rearrangement. The notion of a sheer passive lateral compression of the nucleus might seem questionable in view of the common observation that other spherical cell inclusions, such as fat globules, remain spherical; however, this difference could be accounted for by the difference in consistency between nucleus and fat globules, the former being gelated (to judge from motion pictures of fibroblast cultures), the latter liquid. Even though the meaning of the peculiar constancy of the value of $E_c:E_n$ throughout the range of shapes observed remains obscure, the fact as such is noteworthy in connection with any future attempts to explain nuclear morphology.

Rate of Cell Movement.—In deriving elongate shape from the disparity in the rates of movement of the front and rear poles of the cell, the preceding chapter has already brought out the close relation between cell morphology and cell movement. As the cell elongates, the elastic strain within its cytoplasm (presumably in its fibrous framework) may be assumed to increase proportionately. The adhesive resistance of the rear end will be overcome when this strain has reached the proper magnitude, that is, when a certain critical length has been attained, and this will then be the shape in which we shall find the cell as a whole in locomotion. Just how the gliding type of locomotion characteristic of these cells (and similarly of nerve fiber tips) is achieved, is still rather obscure. It has been compared to the locomotor mechanism of amoebae (W. H. Lewis, 1950), but even the explanations offered for the latter are not yet fully satisfactory. For our objects, both the motor mechanism and the source of the motile energy remain to be revealed. At present, the most appropriate assumption seems to be a dual source, lying partly in the interfacial tensions, partly in intracellular activities supported by metabolic energy. For the nerve fiber, such a dual "pull-push" mechanism of elongation has been suggested previously by one of us (see Weiss, 1944b), and pending proof to the contrary, the fibroblast may be regarded in the same light. That is, we may assume that interfacial tension would provide the "lead force," and that metabolic energy, through some unknown intracellular mechanism would sustain the translatory movement.

On this assumption, the rate of gliding should be mainly a function of the metabolic state of the cell, rather than of the size of the fibers serving as substratum. Elongation and locomotion thus do not necessarily follow the same rules.

Rate of migration was determined by tracing cells at the margin of a culture (always well inside the clot) after 24 and 48 hours of cultivation

and determining their average hourly progress. Figure 8 gives the average rates obtained for 1582 cells so measured in graded plasma concentrations, as well as the values for the 10 fastest cells in each series. Both curves are identical in character. Their extrapolated origin at the null point is consistent with the fact that fibroblasts inside a liquid drop (0% plasma concentration) cannot move at all. The rate of progress then increases very rapidly with increasing plasma concentration, reaching a maximum of about 1 mm. per day near a plasma concentration of 40%, and retaining this value rather constantly throughout the higher concentration range.

In interpreting these curves, it is first to be noted that, as we indicated before, degree of elongation and rate of migration follow different rules, as the former continues to rise (Fig. 7), even after the latter has leveled off. Moreover, since the increase of fibrin fiber size (pertinently reflected in progressive elongation) continues steadily through the whole concentration range, the fact that migratory rate in the upper range remains constant confirms our earlier supposition (see above) that locomotor rate would be independent of fiber size. If this is the case, then the drop of the curves of migratory rate in the lower concentration range must be ascribed to quite different causes than that of the elongation curves, a fact which is further accentuated by the opposite sense of the respective curvatures.

The slower migration in more dilute plasma can evidently be explained on the following grounds. The main clue is the realization that the data of figure 8 were obtained not by a continuous recording of the velocity of cell movement, but simply by measuring the total displacement of cells at the end of a 24-hour period. This would give us a true measure of actual velocity of motion only if cells were moving continuously and at a constant rate throughout the measured interval. However, if the movement is discontinuous, that is, if it is interrupted by stationary phases of unknown duration, slower progress would merely indicate more or longer stops without necessarily implying a slower basic rate. The incidence of such arrests is much higher in the lower plasma concentrations because of the greater frequency of divergent cell processes (Fig. 1*a–c*), as explained earlier in this paper. It is obvious that no cell can advance in two different directions at the same time. Therefore, as long as such processes stalemate each other, the cell as a whole is stalled. Sooner or later one or the other of the competing projections wins out in the resulting tug-of-war, the losing one is withdrawn and the cell can then proceed until the next bifurcation arises, producing another delay, and so on. Our curves, therefore, express cumulative delays rather than retardations of protoplasmic flow.

Another factor vitiating our "rate" curves is the irregularity of the course taken by a cell in any but a strictly oriented medium. Distances in our cultures (Fig. 8) were measured as straight-line connections between the

end points of a 24-hour trek. Actually, however, all multipolar cells had zigzagged about this average course, hence, run a greater distance than that measured, in proportion to the random divergence of their lead processes. Therefore, on this score, too, our "rate" values definitely underrate locomotor speed, the more so the greater the average number of divergent cell processes, that is, the lower the plasma concentration.

Bipolar cells, by contrast, present a wholly different situation, if we remember that bipolarity itself is an expression of gliding on relatively heavy and unintersected path fibers. Having a straight runway, such cells are not subject to the periodical delays of highly intersected networks, for even if they do project more than one advancing process, the fact that all such multiple extensions lie in the common line of advance (Fig. 1-*d*), precludes competitive delays.

These considerations lead to the conclusion that the curves of figure 8 are composed of two separate portions, one above, and the other below, approximately the 40% plasma concentration mark. The former would represent essentially the bipolar cells, which predominate in the higher concentrations (Fig. 5), revealing a basic true gliding velocity of the order of 1 mm. per day independent of fiber size; while the latter would reflect the increasing proportion of multipolar cells in progressively lower concentrations with correspondingly more frequent delays and deviations. It is not unlikely that the actual rate of protoplasmic motion is really the same in both groups. The 1-mm. displacement between 24 and 48 hours of cultivation is in good accord with the maximum rate recorded in a preceding investigation (Weiss, 1945) for the second day of strictly oriented migration of Schwann cells along glass fibers embedded in plasma clots. But our data do not corroborate Willmer's (1933) contention that migration is slower in denser plasma clots. Since neither the experimental conditions nor the methods of measurement nor the sizes of the samples used in his and our experiments were comparable, it is impossible to determine the source of this discrepancy. On the other hand, results quite comparable to ours were obtained in amoebae (Pitts, 1933), in which the

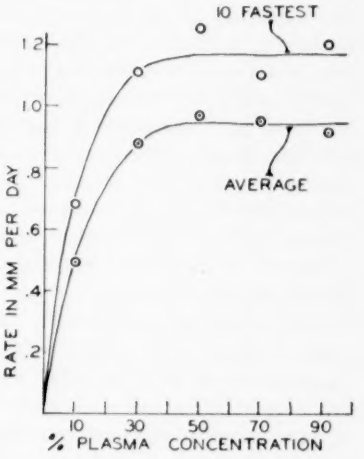

FIGURE 8

Average and fastest migration rates over a 24-hour interval in clots of different textures.

rate of locomotor progress was found to be much slower during multipolar ("multipodal") phases than when they were in the bipolar ("monopodal") form.

Conclusions.—We have demonstrated, in the foregoing, the feasibility of resolving the morphology of a single cell type—the heart fibroblast—quantitatively into terms of its morphogenesis. Pictorial description has been replaced by a number of curves, representing the whole variety of observed cell forms as function of a single variable—the physical structure of the medium. The trend of these curves, at the same time, has verified the basic assumption that the shape of cells of this kind can be derived from an understanding of the manner in which they react to fibrous substrata. It is significant to note that this deviation rests on statistical considerations, reemphasizing the thesis that morphogenesis, by its very nature, is largely statistical. Neither are the individual cells even of a pure cell strain all exactly identical, nor is any organic medium ideally uniform and homogeneous. All that is predictable, therefore, is the average constitution of the cells and the average structure of the medium, hence, the probability of observing a given morphogenetic response within a given range. To this way of thinking, rigid and micro-precise determinism of cell shape is mere illusion. We offer this demonstration of how static morphological symbols can be made to yield to dynamic-statistical interpretations, as encouragement to apply a similar analysis to other cell types.

Such comparative studies of different cell types would have another point in their favor. They could open a wholly new way of getting information about certain properties of living cells uncomplicated by operative tampering. Let us illustrate this by an example. Remember that our data have characterized one single cell type only—the heart fibroblast—in a graded series of environments. All events recorded—the projection of cytoplasmic processes; competitive interaction among multiple processes; rate of advance; elongation; surface reorientation; adhesion to, and detachment from, substratum; etc.—have been predicated on the particular physical and chemical properties of that particular cell type. Cell shapes are no more "caused" or "determined" by the environment as such than they are by any autonomous intracellular factors as such. Cell and environment are two interdependent variables of which we merely kept the one—cell constitution—practically constant throughout the tested range. Now, with this set of data on hand, we can reverse the procedure and go on to comparing different cell types, keeping the environment constant. We can take, for instance, other kinds of mesenchyme cells, nerve sheath cells, sarcoma cells, and so on, determine their parameters (number of processes, elongation of cell body, nuclear deformation, rate of locomotion, etc.) in media of given plasma concentrations and pH

and compare these then with the values obtained from our standard fibroblast strain under corresponding conditions. Any consistent differences can then be taken as indicators of physical or chemical differences between the respective cell strains. For instance, an upward displacement of the curves of figures 5 and 6 (signifying reduced frequency of cell projections) would indicate greater surface stability or swifter protoplasmic movement; an upward shift of the curves in figure 7 (signifying greater elongation), lower viscosity or a lower modulus of elasticity or greater adhesiveness; an increase in the ordinate ratio between the upper and the lower curve of figure 7 (signifying reduced relative elongation of the nucleus), greater turgor or rigidity of the nucleus or reduced strength of the deforming forces in the cytoplasm; etc.—all of these for the particular test strain with reference to our standard strain. Similarly, it would be possible to detect within a given cell strain subtle alterations of physical properties attending differentiation, aging, somatic mutations or pathologic changes.

On the purely technical side, our results illustrate plainly the dangers inherent in the uncritical use of the spread of tissue cultures as a measure of growth. Considering that the increase in area is primarily a result of cell migration, and bearing in mind the complex dependency of the rate and amount of migration upon physical properties of both the cell body and the medium, it is plain that whenever an agent, drug, radiation, etc., has been observed to affect the rate of expansion, this cannot be safely interpreted as an effect on growth as such unless the agent in question has been proved to affect neither the texture of the medium nor such physical conditions in the cell as are relevant to its mode and rate of spreading. The interdependence between cell configuration and cell growth pointed out on a previous occasion (Weiss, 1949) calls for additional caution in the assaying of so-called growth stimulants and growth inhibitors by geometric measurements on tissue cultures.

Summary.—The various cell shapes exhibited by a sample cell type—the heart fibroblast in tissue culture—have been expressed on a single quantitative scale and explained in terms of the reaction of the cell to the fibrous texture of the medium. The texture of the fibrin reticulum was varied systematically in two graded series, one of plasma concentrations, and the other of pH values at coagulation. Measurable criteria of cell shape (number of cell processes; frequency of bipolars in the population; average ratio of length over width of cell body; nuclear elongation) were found to vary accordingly and in the predicted sense. Rate of migration could likewise be related to plasma texture. It has thus been possible, at least for this object, to resolve in principle static cell morphology into quantitative terms of the dynamics of cell morphogenesis.

* Research aided by the Wallace C. and Clara A. Abbott Memorial Fund of the University of Chicago and a grant-in-aid from the American Cancer Society upon the recommendation of the Committee on Growth of the National Research Council.

Hancox, N. M., *Biol. Rev.*, **24,** 448–471 (1949).
Hawn, C. v. Z., and Porter, K. R., *J. Exp. Med.*, **86,** 285–292 (1947).
Lewis, F. T., *Anat. Rec.*, **55,** 323–341 (1933a).
Lewis, F. T., *Proc. Am. Acad. Arts & Sci.*, **68,** 251–284 (1933b).
Lewis, W. H., in *Genetic Neurology* (P. Weiss, ed.), University of Chicago Press, 1950, pp. 53–65.
Matzke, E. B., *Am. J. Bot.*, **34,** 182–195 (1947).
Pitts, R. F., *Biol. Bull.*, **64,** 418–423 (1933).
Porter, K. R., Claude, A., and Fullam, E. F., *J. Exp. Med.*, **81,** 233–246 (1946).
Schmidt, W. J., *Die Bausteine des Tierkörpers im polarisierten Licht*, Bonn, 1924.
Thompson, d'Arcy W., *On Growth and Form*, Cambridge University Press, 1942, 1116 pp.
Weiss, P., *Roux' Arch.*, **116,** 438–554 (1929).
Weiss, P., *J. Exp. Zool.*, **68,** 393–448 (1934).
Weiss, P., *Growth*, **5,** (suppl.) 163–203 (1941).
Weiss, P., *Anat. Rec.*, **88,** 205–221 (1944a).
Weiss, P., *J. Neurosurg.*, **1,** 400–450 (1944b).
Weiss, P., *J. Exp. Zool.*, **100,** 353–386 (1945).
Weiss, P., in *Chemistry and Physiology of Growth* (A. K. Parpart, ed.), Princeton University Press, 1949, pp. 135–186.
Weiss, P., *Quart Rev. Biol.*, **25,** 177–198 (1950).
Weiss, P., and Wang, H., *Proc. Soc. Exp. Biol. & Med.*, **58,** 273–275 (1945).
Willmer, E. N., *J. Exp. Biol.*, **10,** 317–322 (1933).

ERRATA

In my article, "A New Discussion of the Changes in the Earth's Rate of Rotation," in the January issue on page 9 in line 13 from the bottom *7 miles* should read *2.5 miles* and in line 10 from the bottom *200 miles* should read *80 miles*.

DIRK BROUWER

INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. The PROCEEDINGS will aim especially to secure prompt publication of original announcements of discoveries and wide circulation of the results of American research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of others and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of the PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

REFERENCES to literature, numbered consecutively, will be placed at the end of the article and short footnotes should be avoided. It is suggested that references to periodicals be furnished in some detail and in general in accordance with the standard adopted for the Subject Catalogue of the International Catalog of Scientific Literature, viz., name of author, with initials following (ordinarily omitting title of paper), abbreviated name of Journal, with place of publication, series (if any), volume, inclusive pages, year. For example: Montgomery, T. H., *J. Morph., Boston*, 22, 731–815 (1911); or, Wheeler, W. M., *Kewigsberg, Schr. physik. Ges.*, 25, 1–142 (1914).

MANUSCRIPTS will be accepted only from members of the Academy or from chairmen of the divisions of the National Research Council who will assume responsibility for the propriety and scientific standards of the paper and should be addressed to *Proceedings of the* NATIONAL ACADEMY OF SCIENCES, *2101 Constitution Ave., Washington, 25, D. C. Articles of which a member of the Academy is author or a co-author may extend to twelve printed pages, those by non-members only to six printed pages; space above these limits will be billed to the author at $17.50 per page.*

PROOF will ordinarily be sent,—with the understanding that charges for an author's correction shall be billed to him. Authors are therefore requested to make final revisions on the typewritten manuscripts.

REPRINTS should be ordered at the time of submission of manuscript. They will be furnished to authors at cost, approximately as follows:

SCHEDULE OF RATES FOR REPRINTS, CARRIAGE CHARGES EXTRA

No. Copies	50	100	150	200	250
4 pp.	4.85	5.20	5.85	6.50	7.15
8 pp.	9.10	10.40	11.70	13.00	14.30
Covers	3.45	4.48	5.41	6.50	7.57

CONTENTS

Page

CHEMISTRY.—CHEMICAL VALENCE FORCES AND BINDING ENERGY CALCULATIONS . . . . . *By Charles P. Mueller and Henry Eyring* 149

PHYSICS.—SUPERPOSITION OF CONFIGURATIONS: THE HELIUM ATOM . . . . . *By G. Russell Taylor and Robert C. Parr* 154

PHYSICS.—A COMPARATIVE SURVEY OF APPROXIMATE GROUND STATE WAVE FUNCTIONS OF HELIUM ATOM AND HYDROGEN MOLECULE . . . . . *By R. S. Mulliken* 160

BIOCHEMISTRY.—THE RADIATION SENSITIVITY OF CATALASE AS A FUNCTION OF TEMPERATURE . . . . . *By R. B. Setlow* 166

BIOCHEMISTRY.—THE PHYSICAL INACTIVATION OF INVERTASE . . . . . *By Ernest Pollard, W. F. Powell and S. H. Reaume* 173

BIOCHEMISTRY.—IONIZING PARTICLE EVIDENCE FOR THE MOLECULAR WEIGHT OF THE PNEUMOCOCCUS TRANSFORMING PRINCIPLE . . . . . *By Donald Fluke, Ruth Drew and Ernest Pollard* 180

BIOCHEMISTRY.—STUDIES ON THE HEMOGLOBIN OF COOLEY'S ANEMIA AND COOLEY'S TRAIT . . . . . *By Alexander Rich* 187

GENETICS.—THRESONINE INHIBITION IN A STRAIN OF NEUROSPORA . . . . . *By C. O. Doudney and R. P. Wagner* 196

GENETICS.—OBSERVATIONS ON THE BEHAVIOR OF SUPPRESSORS IN NEUROSPORA . . . . . *By Mary B. Mitchell and Herschel K. Mitchell* 205

GENETICS.—THE EFFECTS OF GENE CHANGE ON TRYPTOPHAN DESMOLASE FORMATION . . . . . *By Charles Yanofsky* 215

MATHEMATICS.—PARTIAL DIFFERENTIAL EQUATIONS AND ANALYTIC CONTINUATIONS . . . . . *By S. Bochner* 227

MATHEMATICS.—THE DIRICHLET PROBLEM FOR LINEAR ELLIPTIC EQUATIONS OF ARBITRARY EVEN ORDER WITH VARIABLE COEFFICIENTS . . . . . *By Felix E. Browder* 230

MATHEMATICS.—INTEGRATION OF STIFF EQUATIONS . . . . . *By C. F. Curtiss and J. O. Hirschfelder* 235

MATHEMATICS.—THE ELEMENTARY BASIC PRINCIPLES OF THE UNIFIED THEORY OF RELATIVITY . . . . . *By Václav Hlavatý* 243

MATHEMATICS.—HOMOLOGY RELATIONS ON REGULAR ORIENTABLE MANIFOLDS . . . . . *By Marston Morse* 247

MATHEMATICS.—CRITERIA FOR LINEAR EQUIVALENCE . . . . . *By André Weil* 258

PATHOLOGY.—RAPID EFFECTS UPON THE RENAL CIRCULATION PRODUCED BY NEPHROTOXIC GLOBULIN ADMINISTRATION IN THE RAT . . . . . *By Richard W. Lippman, Helen U. Marti and E. Elmo Jacobs* 260

ZOOLOGY.—SHAPE AND MOVEMENT OF MESENCHYME CELLS AS FUNCTIONS OF THE PHYSICAL STRUCTURE OF THE MEDIUM. CONTRIBUTIONS TO A QUANTITATIVE MORPHOLOGY . . . . . *By Paul Weiss and Beatrice Garber* 264

ERRATA.—A NEW DISCUSSION OF THE CHANGES IN THE EARTH'S RATE OF ROTATION . . . . . *By Dirk Brouwer* 288